



Un guide virtuel autonome pour la description d'un environnement réel dynamique: interaction entre la perception et la prise de décision

Morgan Veyret

► To cite this version:

Morgan Veyret. Un guide virtuel autonome pour la description d'un environnement réel dynamique: interaction entre la perception et la prise de décision. Autre [cs.OH]. Université de Bretagne occidentale - Brest, 2009. Français. NNT: . tel-00376176

HAL Id: tel-00376176

<https://theses.hal.science/tel-00376176>

Submitted on 16 Apr 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ DE BRETAGNE OCCIDENTALE

— Mémoire de thèse —

Spécialité : Informatique

Un guide virtuel autonome pour la description d'un
environnement réel dynamique
Interaction entre la perception et la prise de décision

MORGAN VEYRET

Pour l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE BRETAGNE OCCIDENTALE
Mention INFORMATIQUE
LISYC – *Laboratoire d'Informatique des Systèmes Complexes* – EA3883

Soutenue publiquement le 6 mars 2009 devant la commission d'examen :

Rapporteurs

Jean-Pierre	JESSEL	Professeur à l'Université de Toulouse III
Malik	MALLEM	Professeur à l'Université d'Evry val d'Essonne

Examineurs

Patrick	REIGNIER	Maître de conférences à l'Université de Grenoble
Vincent	RODIN	Professeur à l'UBO (Président)
Eric	MAISEL	Maître de conférences à l'ENIB
Jacques	TISSEAU	Professeur à l'ENIB

Invité

Sylvain	GHIRON	Responsable multimédia à Océanopolis
---------	--------	--------------------------------------

Un guide virtuel autonome pour la description d'un environnement réel dynamique

Interaction entre la perception et la prise de décision

Mémoire de thèse

MORGAN VEYRET

— 1^{er} avril 2009 —

Laboratoire d'Informatique des SYstèmes Complexes EA 3883 UBO-ENIB

MORGAN VEYRET

e-mail : veyret@enib.fr

url : [http ://www.enib.fr/~veyret](http://www.enib.fr/~veyret)

tel : +33 (0)2 98 05 66 66



UBO Université de Bretagne Occidentale

[http ://www.univ-brest.fr](http://www.univ-brest.fr)



CERV/ENIB Centre Européen de Réalité Virtuelle Ecole Nationale d'Ingénieurs de Brest

Technopôle Brest-Iroise

25, rue Claude Chappe

BP 38 F-29280 Plouzané

[http ://www.cerv.fr](http://www.cerv.fr)

[http ://www.enib.fr](http://www.enib.fr)



Région Bretagne

[http ://www.region-bretagne.fr](http://www.region-bretagne.fr)



Remerciements

Je remercie avant tout l'équipe du CERV pour son accueil et sa sympathie durant ces années de thèse. Je tiens à remercier Jacques TISSEAU pour la confiance qu'il m'a témoigné en m'accueillant dans son laboratoire et en acceptant la direction de cette thèse ainsi qu'Éric MAISEL pour m'avoir encadré.

Je remercie Jean-Pierre JESSEL et Malik MALLEM pour avoir accepté de rapporter cette thèse. J'adresse également mes remerciements à Patrick REIGNER et Vincent RODIN pour avoir participé à mon jury.

Ces travaux n'auraient pu aboutir sans le soutien de nombreuses autres personnes et notamment Stéphane, Cyrille, les fantômes de `#glute` et ceux que j'oublie.

Table des matières

Remerciements	iii
Table des matières	v
Table des figures	xi
Liste des tableaux	xv
Liste des listings	xvii

partie I	— Introduction —	1
-----------------	-------------------------	----------

1 Introduction	3
1.1 Contexte	3
1.1.1 Présentation	3
1.1.2 Évolution des moyens de communication avec le visiteur	4
1.1.3 Réalité augmentée	6
1.1.4 Agent autonome et acteur virtuel	8
1.2 Problématique et proposition	10
1.3 Organisation du mémoire	12

partie II	— Décrire le réel —	13
------------------	----------------------------	-----------

2 Du guide réel au guide virtuel autonome	15
2.1 Introduction	15
2.2 Visite guidée et guide réel	15
2.3 Visite guidée et guide virtuel	18
2.3.1 Notion de guide virtuel	18
2.3.2 La visite : la construction d'un exposé	20
2.3.3 Bilan	23
2.4 Structuration de la visite	24

2.4.1	Visite statique	24
2.4.2	Visite interactive	25
2.4.3	Visite réactive	26
2.4.4	Visite dynamique	27
2.4.5	Bilan	30
2.5	Expression et communication	31
2.5.1	Incarnation et personnification	31
2.5.2	Communication multimodale	32
2.5.3	Incarnation et comportement	34
2.5.4	Bilan	36
2.6	L'adaptation	37
2.6.1	Personnalisation ou adaptivité	37
2.6.2	Objets de l'adaptation	37
2.6.3	Sources d'adaptation	38
2.6.4	Bilan	40
2.7	Conclusion	40
3	Un guide virtuel autonome pour un environnement dynamique	43
3.1	Introduction	43
3.2	Les explications	44
3.2.1	Introduction	44
3.2.2	La notion d'explication	45
3.2.3	Graphe d'explication	46
3.2.4	Exécution : processus d'explication	47
3.2.5	Transitions entre explications	49
3.2.6	Bilan	50
3.3	La visite guidée : une succession d'explications	51
3.3.1	Introduction	51
3.3.2	Notion de sujet	51
3.3.3	Contexte de la visite	52
3.3.4	Sélection des explications	55
3.3.5	La visite guidée	57
3.3.6	Bilan	58
3.4	Comportement du guide virtuel	58
3.4.1	Introduction	58
3.4.2	Notion de comportement	59
3.4.3	Exécution et gestion de l'interruptibilité	61
3.4.4	Bilan	63
3.5	Déroulement d'une visite	64
3.6	Conclusion	65
	 partie III — Percevoir le réel —	 67
4	Représenter le réel pour le décrire	69
4.1	Introduction	69
4.2	Contexte et contraintes	69
4.3	Perception visuelle	70

4.3.1	Une vue élémentaire ou globale ?	70
4.3.2	Perception directe ou inférence ?	71
4.3.3	Percevoir en fonction de la tâche	72
4.3.4	Mémoire et représentation	73
4.4	Approches de la vision artificielle	75
4.4.1	Vision active	75
4.4.2	Perception active	76
4.4.3	Vision animée	76
4.4.4	Vision intentionnelle	76
4.5	Vision artificielle	77
4.5.1	Vision par ordinateur	77
4.5.2	Les systèmes de vision : du traitement à l'interprétation	78
4.5.3	Représentation des informations	79
4.5.4	Architectures logicielles et coopération	81
4.5.5	Le contrôle	83
4.5.6	Bilan	86
4.6	Attention visuelle	87
4.6.1	Vision précoce ou pré-attentive	88
4.6.2	Traitements séquentiels	91
4.6.3	L'influence du contexte	93
4.6.4	Attention et représentation	96
4.6.5	Contrôle attentionnel	98
4.6.6	Bilan	101
4.7	Conclusion	102
5	Le contrôle de l'attention visuelle pour construire une représentation	105
5.1	Introduction	105
5.2	Une représentation partielle de l'environnement	106
5.3	Organisation du système de perception	108
5.3.1	Introduction	108
5.3.2	Mémoire visuelle	108
5.3.3	Extraction d'informations à partir des images	112
5.3.4	Routines de perception	115
5.3.5	Bilan	117
5.4	Construction d'une représentation : exécution et contrôle	117
5.4.1	Introduction	117
5.4.2	Stratégies de perception	118
5.4.3	Influence du processus de prise de décision : notion d'intérêt	120
5.4.4	Contrôle de l'intérêt : la représentation comme interface	122
5.4.5	Précisions sur l'implémentation	125
5.4.6	Bilan	125
5.5	Percevoir pour décrire l'environnement	126
5.5.1	Introduction	126
5.5.2	Organisation pour la description d'un environnement dynamique	127
5.5.3	Stratégies de perception	130
5.5.4	Contrôle du processus de prise de décision	136
5.5.5	Bilan	139

5.6	Conclusion	140
partie IV — Application et expérimentations —		143
6	Application et expérimentations	145
6.1	Introduction	145
6.2	Application : le projet SIRENE	145
6.2.1	Besoins en terme de présentation d'informations	145
6.2.2	Mise en œuvre	147
6.2.3	Comportement et discours	148
6.2.4	Perception visuelle	154
6.2.5	Bilan	158
6.3	Expérimentations : la visite guidée	159
6.3.1	Introduction	159
6.3.2	Gestion des interruptions	159
6.3.3	Sélection des explications	162
6.3.4	Bilan	170
6.4	Expérimentations : perception visuelle	171
6.4.1	Introduction	171
6.4.2	Représentation de l'environnement	171
6.4.3	Contrôle attentionnel	175
6.4.4	Bilan	181
6.5	Conclusion	182
partie V — Bilan et perspectives —		183
7	Conclusion générale	185
7.1	Bilan	185
7.2	Perspectives	188
Références bibliographiques		193
A	Définition de la visite guidée	213
A.1	Introduction	213
A.2	Définition des éléments de discours	213
A.3	Sujets, explications et transitions	214
A.3.1	Définition d'une explication	215
A.3.2	Définition d'un sujet	215
A.3.3	Définition d'une transition	215
A.3.4	Prototypage	216
A.4	Définition du comportement	217
A.5	Définition et activation des experts	219
B	Rendu et visualisation	221
B.1	Introduction	221
B.2	Protocole de contrôle	222
B.2.1	Commandes générales	222

B.2.2	Contrôle de l'état du guide virtuel	222
B.2.3	Navigation	223
B.2.4	Animations et médias	223
B.3	Gestion de la navigation	224
C	Implémentation du système de perception visuelle	227
C.1	Architecture logicielle	227
C.1.1	Un ensemble d'objets actifs qui coopèrent	228
C.1.2	Exécution et communication	229
C.2	Des objets actifs au système de perception visuelle	230
C.2.1	Comportements et routines de perception	231
C.2.2	Traitement d'image	231
C.2.3	Mémoire visuelle	232
C.2.4	Notes sur la distribution	232
D	Reconnaissance et classification des espèces de poisson	235
D.1	Classifieurs et <i>features</i>	235
D.2	Reconnaissance et apprentissage	236

Table des figures

— Introduction —	1
1.1 Exemples de monde augmenté par le système Archeoguide et vue du système sur un utilisateur [Vlahakis 02].	6
1.2 Photo du robot-guide MINERVA [Thrun 99].	7
1.3 Le continuum Réel/Virtuel [Milgram 94]	7
1.4 Vue du système de réalité augmentée ARQuake [Thomas 02] et vue de l'utilisateur.	8
1.5 Illustration des problématiques classiquement associées aux systèmes de réalité augmentée (de haut en bas et de gauche à droite) : incrustation arbitraire, prise en compte des problèmes d'alignement, prise en compte de la cohérence spatio-temporelle et enfin prise en compte de la cohérence photométrique. Images extraites de http ://www.loria.fr/~gsimon/ra/	9
1.6 Vue d'ensemble de notre proposition.	11
— Décrire le réel —	13
2.1 Exemples de guides mobiles	18
2.2 Exemples de robots guides utilisés en condition réelles au sein d'un musée [Graf 02].	19
2.3 Exemples de guides virtuels en environnement virtuel	20
2.4 Exemple de guide virtuel faisant appel à la réalité augmentée en extérieur [Braun 03]	21
2.5 Facilité de créer des plans <i>a priori</i> pour la génération automatique de discours [Tanaka-Ishii 00].	24
2.6 Vue d'ensemble du discours organisé sous forme de schéma [Yuan 03].	28
2.7 Architecture globale du système VITRA [Herzog 94].	29
2.8 Capture d'écran de l'interface du système GUIDE [Cheverst 00].	33
2.9 L'incarnation du guide d'accueil présenté dans [Kopp 05].	34
2.10 Comportement sous forme de machine à état interruptible dans [Klesen 03].	35
3.1 Positionnement du chapitre dans l'architecture générale de notre proposition.	44

3.2	Exemple de graphe d'explication.	47
3.3	Vue d'ensemble des données de discours et d'une instance de visite guidée (chemin au sein de ces données).	57
3.4	Exemple d'automate hiérarchique pouvant être utilisé dans le cadre du guide virtuel.	60
3.5	Exemples de sous-automates interruptibles	61
 — Percevoir le réel —		 67
4.1	Modèle de mémoire proposé par [Atkinson 68]	73
4.2	La mémoire de travail proposée par [Baddeley 74]	74
4.3	Différents types de représentation.	81
4.4	Différentes formes de coopération entre sources d'informations [Garbay 00]. (a) coopération confrontative; (b) coopération augmentative; (c) coopération intégrative.	83
4.5	Modèle de carte de saillance proposé par Itti & Koch [Itti 98].	89
4.6	Exemple de carte de saillance	89
4.7	Parcours de l'œil dans une image [Yarbus 67].	91
4.8	Parcours de l'œil dans une image et ses variations en fonction de la tâche [Yarbus 67]. Dans chaque trace, les sujets devaient examiner l'image : 1 – sans but particulier ; 2 – pour estimer le niveau de vie des personnages ; 3 – pour estimer l'âge des personnages ; 4 – pour deviner ce que faisaient les personnages avant l'arrivée des visiteurs ; 5 – mémoriser les vêtements des personnages ; 6 – mémoriser la position des objets et personnages dans la pièce ; 7 – estimer le temps depuis la dernière visite des invités.	94
4.9	Image montrant le rôle de l'interprétation dans la perception visuelle.	94
4.10	Le modèle attentionnel multifocal [Cavanagh 05].	98
4.11	Représentation des <i>object files</i> [Kahneman 92; Wolfe 97].	99
4.12	Organisation des routines visuelles dans [Salgian 98a].	100
4.13	Contextualisation et exécution des routines visuelles dans [Herviou 06].	101
5.1	Positionnement de ce chapitre dans l'architecture générale de notre proposition.	105
5.2	Mémoire visuelle. L'environnement y est représenté comme un ensemble d'entités visuelles décrites par des propriétés.	109
5.3	Les différents états possibles pour une entité visuelle.	110
5.4	Paramètres de gestion de la durée de vie d'une entité visuelle.	110
5.5	Évolution du nombre d'entités dans chacun des états possible : provisoire (rouge) ; valide (vert) ; active (bleu).	112
5.6	Exemple d'utilisation de cartes de propriétés.	114
5.7	Résultat de l'exécution de routines visuelles.	116
5.8	Organisation des traitements dans le système de perception visuelle.	116
5.9	Exécution et contrôle au sein du système de perception visuelle.	118
5.10	Boucle entre traitements et contenu de la mémoire lors de l'exécution des routines de perception.	119
5.11	Vue schématique du mécanisme de contrôle par les requêtes sur la représentation de l'environnement.	124
5.12	Organisation du système de perception visuelle dans le cadre du guide virtuel.	128

5.13	Les différentes « couches » du mécanisme de reconnaissance.	129
5.14	Exemples de variations du délai courant des routines de détection en fonction des valeurs d'intérêt.	132
5.15	Capture d'écran du suivi de cibles au sein de notre application à un aquarium marin.	133
5.16	Organisation des routines pour le comportement de reconnaissance en processus de classification.	135
5.17	Variations de la priorité des routines de reconnaissance en fonction de l'intérêt porté sur les entités et les types. Chaque plan correspond à un couple possédant un état d'avancement particulier.	137
5.18	Représentation graphique des différentes contraintes pouvant être utilisées lors des requêtes sur la représentation de l'environnement.	138
 — Application et expérimentations —		143
6.1	Des visiteurs face à l'un des bassins d'Océanopolis.	146
6.2	Exemples de fiches de présentation de différents poissons d'un des bassins du pavillon tropical d'Océanopolis.	147
6.3	Le bassin du point de vue du visiteur au travers du système de visualisation et l'utilisateur devant le prototype du système.	148
6.4	Démonstration dans le cadre de la Nuit des Chercheurs du 26 Septembre 2008.	149
6.5	Comportement global du guide virtuel dans le cadre de l'application.	150
6.6	Exemple de comportement d'attente.	151
6.7	Comportement d'explication directe avec entité visuelle.	152
6.8	Comportement d'explication avec entité visuelle comportant une phase de transition.	153
6.9	Capture d'écran du contenu de la représentation de l'environnement dans l'application au bassin d'Océanopolis.	155
6.10	Les trois espèces considérées dans le cadre du projet SIRENE.	156
6.11	Classifieurs et propriétés pour la reconnaissance des espèces Platax, Naso et Requin.	157
6.12	Capture d'écran de la vidéo utilisée pour les expérimentations sur le système de perception visuelle.	172
6.13	Nombre de cibles suivies et reconnues pour des limitations à 10, 15 et 20 images par secondes pour l'objet effectuant les traitements.	173
6.14	Nombre de cibles suivies et disparues/perdus au cours du temps pour des limitations de 10, 15 et 20 images par seconde.	174
6.15	Nombre de cibles reconnues pour chacun des types possibles (red , green et blue) pour des limitations à 10, 15 et 20 images par secondes.	175
6.16	Paramètres du système sous influence du mécanisme de contrôle avec une limitation à 10 images par seconde.	177
6.17	Paramètres du système sous influence du mécanisme de contrôle avec une limitation à 15 images par seconde.	178
6.18	Nombre d'entités de chaque type sans et avec contrôle attentionnel pour une limitation à 15 images par secondes.	179
6.19	Nombre d'entités de chaque type sans et avec contrôle attentionnel pour une limitation à 15 images par secondes.	180

6.20	Évolution de l'intérêt au sein du comportement de vigilance pour différents taux d'atténuation (5%, 10% et 50% par seconde).	181
 — Bilan et perspectives —		 183
7.1	Vue d'ensemble de l'architecture proposée.	186
B.1	Représentation du guide virtuel sous la forme d'un Diodon.	222
B.2	Captures d'écran de la fenêtre du moteur de rendu avec l'affichage des cibles servant à la gestion de la navigation.	225
C.1	Vue d'ensemble de l'architecture de communication et d'exécution d'objets actifs sur laquelle repose le système de perception visuel. Chaque « nuage » représente une plateforme d'exécution d'un ensemble d'objets actifs (les cercles). Chaque plateforme possède un module de communication (carré) se chargeant de la communication par message entre objets et de la recherche de service par un système d'annuaire.	228

Liste des tableaux

— Introduction —	1
— Décrire le réel —	13
2.1 Extrait des documents utilisés à Océanopolis pour diriger la visite guidée du pavillon tropical. La visite totale est prévue pour durer environ 90 minutes. .	17
3.1 Exemples de transitions et situations dans lesquelles elle doivent être utilisées.	49
— Percevoir le réel —	67
— Application et expérimentations —	143
6.1 Liste des sujets disponibles ainsi que les thèmes qui leur sont associés.	153
6.2 Résultats du système de reconnaissance pour les trois espèces de poissons considérées dans le cadre du projet SIRENE.	157
6.3 Déroulement d’une visite sur une évolution de l’environnement déterminée. . .	160
6.4 Évolution de la représentation de l’environnement.	165
6.5 Déroulement de la visite pour différentes configurations d’experts.	166
— Bilan et perspectives —	183

Liste des listings

3.1	Exemple d'interrogation de la représentation de l'environnement	54
5.1	Exemples d'interrogation de la représentation de l'environnement.	122
A.1	Exemple de déclaration d'un élément de discours.	214
A.2	Exemple de déclaration d'une explication (scénario d'explication).	215
A.3	Exemple de déclaration d'un sujet.	215
A.4	Exemple de déclaration d'une transition.	216
A.5	Syntase utilisée pour la définition des données de discours par les auteurs. . .	216
A.6	Exemple de déclaration d'un comportement local et représentation graphique de celui-ci.	218
A.7	Exemple de déclaration d'un expert.	219
C.1	Exemples de services et de leurs propriétés. Ici deux services sont disponibles, l'un de classification et l'autre de mise à jour d'image à partir des caméras. Plusieurs objets actifs peuvent fournir le même service, un service particulier étant décrit par un ensemble de propriétés.	229
C.2	Exemple de message	230

Partie I

— Introduction —

Chapitre 1

Introduction

1.1 Contexte

1.1.1 Présentation

Les travaux présentés dans cette thèse ont été menés au CERV¹ dans le cadre du projet SIRENE² en collaboration avec le musée marin Océanopolis³. Ce projet vise à l'étude de la description d'un *environnement réel dynamique* par un acteur virtuel *autonome* immergé dans cet environnement à l'aide de la *réalité augmentée*.

Les musées tels qu'Océanopolis fournissent des informations aux visiteurs au travers d'expositions. Leur rôle est de montrer aux visiteurs ce qu'il faut regarder et ce qu'il faut apprendre. Pour cela, ils utilisent principalement deux moyens :

- ▷ les expositions elles-mêmes (sélection, arrangement des objets présentés),
- ▷ la fonction de *docent* [Rayward 99] qui consiste en la mise en place de supports de communication avec le visiteur, la formation de guides et la création de visites commentées et/ou d'activités éducatives.

Les musées ne se résument donc pas simplement à une collection d'objets mais tentent d'apporter une valeur ajoutée à cette collection en communiquant aux visiteurs des informations concernant ces objets à travers différents moyens.

Des moyens classiques existent depuis longtemps : simples panneaux d'affichage placés près de l'objet présenté, brochures à suivre tout au long de la visite ou encore visites

¹ Centre Européen de Réalité Virtuelle (<http://www.cerv.fr>).

² Système Ichtyologique de Renseignement Educatif Nautique (<http://www.anr-sirene.com>).

³ Océanopolis : parc de la découverte des océans (<http://www.oceanopolis.com>).

accompagnées d'un guide humain. Les panneaux et brochures sont les plus répandus – les visites guidées nécessitant des moyens humains qui ne sont pas toujours disponibles – ils souffrent néanmoins d'un certain nombre d'inconvénients. L'espace disponible, à la fois sur le support d'affichage et dans le musée pour positionner ces supports, est problématique de ce type de moyens de communication. Dans le cas d'un aquarium marin par exemple, il n'est pas possible au regard du grand nombre d'espèces présentes dans un bassin de toutes les décrire. Le visiteur, lorsqu'il n'assiste pas à une visite guidée, ne dispose que d'informations limitées sur ce qu'il peut observer. Les descriptions proposées sur un panneau explicatif présentent des informations succinctes par manque de place principalement, mais également pour ne pas « assommer » les visiteurs qui cherchent avant tout à se détendre en s'instruisant [Kelly 02]. De plus, les informations présentées sont les mêmes pour tous les visiteurs. Elles sont généralement adaptées à un niveau moyen afin de satisfaire le plus grand nombre de visiteurs.

Nous proposons au travers du projet SIRENE de palier à ces limitations grâce à l'intégration d'un guide virtuel au sein de l'environnement à décrire. L'idée est que ce guide apparaisse aux visiteurs comme faisant partie intégrante de l'environnement qu'il présente, établissant un lien direct entre ses explications et les entités qu'il décrit. Dans la suite de ce chapitre nous présentons le contexte de nos travaux avant de préciser notre problématique. Ce contexte repose sur trois points : les différents moyens mis en œuvre par les musées pour communiquer avec le visiteur, une vue d'ensemble des problèmes de la réalité augmentée et la notion d'acteur virtuel autonome.

1.1.2 Évolution des moyens de communication avec le visiteur

Les nouvelles technologies telles que l'informatique et le multimédia ont très tôt intéressées les musées [Bearman 91] et apportées d'intéressantes perspectives en termes de moyens de communication avec le visiteur tout en ayant un aspect attractif. Un exemple classique de l'utilisation de ces technologies est celui des guides audio et/ou vidéo proposés à l'entrée des musées. Les CD-ROMs et autres kiosques d'information ont apporté la notion d'interactivité dans la transmission de l'information : le visiteur devient actif dans sa quête d'informations. Dans [Heinecke 95], des systèmes de kiosques multimédia sont évalués. Ils se présentent sous la forme d'écrans (tactiles ou non) permettant d'afficher des pages multimédia et de naviguer parmi ces pages à l'aide de boutons et de liens hypertextes. Ces systèmes intriguent les visiteurs et les incitent à consulter les informations qu'ils proposent [Economou 98].

Avec l'arrivée d'Internet, les musées ont trouvé un nouveau moyen de communication important au travers de sites *web* qui se sont peu à peu transformés en véritables musées virtuels⁴ [Bearman 95]. Les limites d'espace et d'adaptation sont abolies mais le visiteur n'est plus en contact direct avec les objets exposés. Cette virtualisation du musée permet d'offrir de nouvelles expositions aux visiteurs mais nécessite la numérisation et le stockage de collections virtuelles [Pietsch 04; Pollefeys 01]. L'absence de contact avec les objets réels de l'exposition peut constituer un problème pour certaines de ces structures telles qu'un aquarium marin. En effet, l'intérêt des expositions proposées par ce type de structure est, pour les visiteurs,

⁴ Le lecteur pourra se référer à [Arc] pour un ensemble de conférences et d'articles sur le rapport entre les musées et le multimédia notamment pour le cas d'Internet.

la possibilité qu'elles offrent de voir « en vrai » des espèces d'animaux du monde entier et la possibilité de les observer dans leur environnement « réel » ⁵.

L'informatique ubiquitaire introduite par [Weiser 91] et d'une façon plus générale l'utilisation de terminaux mobiles interactifs [Deshayes 04] a été une source d'inspiration pour de nombreuses applications dans le cadre des musées. [Gool 99] par exemple propose l'utilisation d'un terminal mobile équipé d'une caméra. Celle-ci permet d'identifier le tableau devant lequel le visiteur se trouve et de fournir des informations spécifiques sur celui-ci. Ces systèmes mobiles ont apporté une idée nouvelle dans les moyens de communication avec le visiteur : l'adaptation au contexte. En effet, les systèmes utilisés jusque-là ne présentaient soit aucune adaptation (moyens « classiques »), soit une adaptation limitée, principalement guidée par les choix du visiteur. L'adaptation par rapport au contexte dans un sens plus général (position, orientation le plus souvent) a permis de dépasser le problème posé par les points d'information multimédia qui se voyaient restreints géographiquement au sein du musée. Désormais, l'information accompagne le visiteur.

La notion de réalité virtuelle a également intéressé les musées en tant que moyen de communication. Soit au travers de musées virtuels copiant un musée réel au sein duquel le visiteur retrouve à la fois les avantages des collections numériques et l'espace du musée (et de ce fait l'exposition au sein du musée physique) [Yuan 03; Lepouras 01]. Soit pour permettre aux visiteurs de découvrir des endroits difficilement accessibles (la planète Mars, une cité disparue, les fonds marins [Dinis 08], *etc.*). L'utilisation d'interfaces immersives et de périphériques haptiques a également été abordée afin d'offrir au visiteur la sensation d'observer des objets réels [Bergamasco 01].

Nombre d'applications ont également fait appel à la réalité augmentée (cf. section 1.1.3) afin de fournir des informations aux visiteurs. Ubiquitous Talker propose, par exemple, l'utilisation d'une caméra et de techniques de réalité augmentée pour fournir des commentaires audio à l'utilisateur lorsqu'il observe un objet [Nagao 95]. Dans [Hughes 04], un système de réalité mixte plonge les visiteurs au sein du monde aquatique du Crétacé à l'aide de lunettes semi-transparentes. L'utilisation de la réalité augmentée permet, au même titre que la réalité virtuelle, de montrer au visiteur des choses impossibles dans le monde réel. Elle permet également de conserver le lien physique entre le visiteur et l'exposition.

Dès 1996, [MacIntyre 96] s'est intéressé à l'utilisation de ces deux types de technologie : la réalité augmentée et l'informatique ubiquitaire (malgré leur coût élevé à l'époque) pariant, à juste titre, sur leur démocratisation. Il voit en la combinaison de ces deux idées, au travers de systèmes mobiles de réalité augmentée, le futur des interfaces utilisateur. Il met en avant la difficulté de la création du contenu dans de telles interfaces et propose l'utilisation d'agents de présentation pour gérer les différentes modalités disponibles dans ces systèmes. Des applications mobiles faisant appel à la réalité augmentée ont alors vues le jour rapidement [Starner 97] et restent encore développées aujourd'hui. C'est le cas du système Archeoguide (figure 1.1) qui fait appel à la réalité augmentée en extérieur afin de recréer le site historique d'Olympie aujourd'hui disparu et d'y faire vivre des acteurs virtuels dans des épreuves olympiques [Vlahakis 02].

⁵ L'aquarium marin d'Océanopolis par exemple s'attache à recréer des écosystèmes adaptés aux espèces présentées.

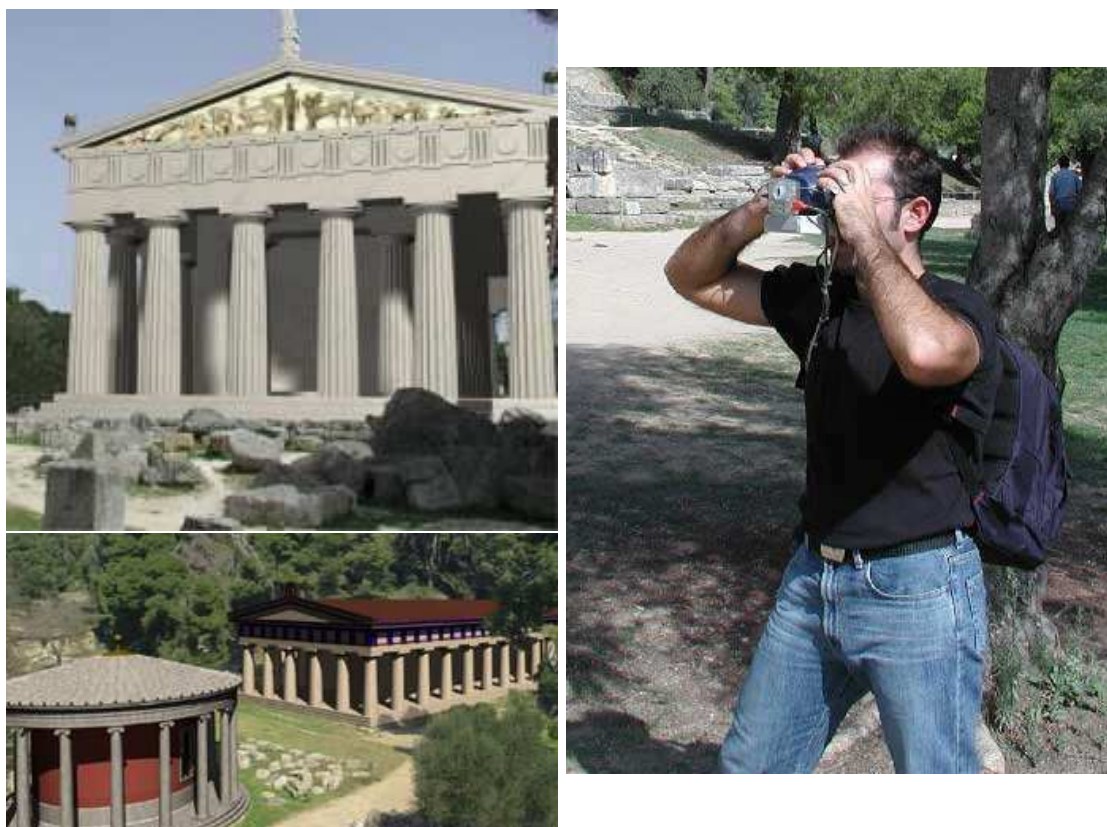


FIG. 1.1 – Exemples de monde augmenté par le système Archeoguide et vue du système sur un utilisateur [Vlahakis 02].

Dans [Burgard 98], un robot autonome (RHINO) est à même de faire visiter à un groupe de personnes le *Deutsches Museum* de Bonn en Allemagne. Sa version améliorée MINERVA [Thrun 99] ajoute un visage au robot afin de lui offrir une plus grande expressivité (figure 1.2). Ce type de moyen de communication met en avant l’incarnation, la présence et l’interaction afin de faciliter la transmission d’informations aux visiteurs.

L’idée du projet SIRENE est de tirer parti de cette notion de présence, d’incarnation du système de présentation d’information tout en conservant et renforçant le lien entre le visiteur et les objets décrits.

1.1.3 Réalité augmentée

Nous avons évoqué la réalité augmentée comme un moyen d’ajouter des informations au monde réel. Nous apportons ici quelques précisions quant à la définition et aux problèmes liés à cette idée.

Selon [Azuma 01] un système de réalité augmentée doit posséder les caractéristiques suivantes :

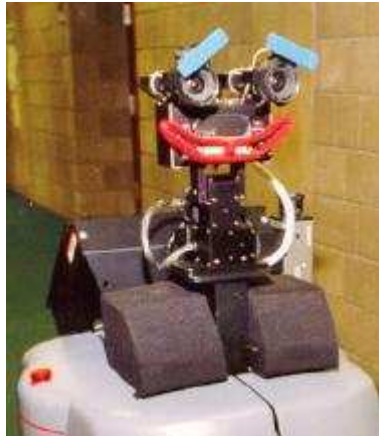


FIG. 1.2 – Photo du robot-guide MINERVA [Thrun 99].

- ▷ combiner des objets réels et virtuels dans un environnement réel ;
- ▷ être interactif et temps réel ;
- ▷ faire appel à une représentation 3D des objets.

Cette définition ne se restreint à aucun système de visualisation (écran vidéo, lunettes semi-transparentes, ...) ni à l'augmentation d'une seule modalité (la modalité visuelle généralement) et inclut de fait une grande variété d'applications. [Milgram 94] désigne sous le terme de *réalité mixte* les techniques qui mettent en œuvre un mélange entre des objets réels et virtuels. Il les organise selon un continuum présenté dans la figure 1.3. La caractéristique principale de la réalité augmentée est son ancrage fort au sein d'un environnement réel contrairement à la *Réalité Virtuelle* ou la *Virtualité Augmentée*⁶ pour lesquelles l'environnement est virtuel.

Le domaine de la réalité augmentée consiste généralement en l'ajout d'informations au monde réel à l'aide de technologies de fusion d'images comme par exemple des écrans ou lunettes semi-transparentes (figure 1.4). Néanmoins, il existe de nombreuses formes d'augmentation de la réalité. Ainsi, il est possible d'enrichir le monde réel suivant différentes modalités, sonores, visuelles, voir olfactives. Il est également possible d'enrichir le monde dans sa globalité ou de se concentrer sur l'enrichissement d'objets particuliers.

La mise en œuvre de systèmes de réalité augmentée pose un ensemble de problèmes qui ne sont encore à ce jour que partiellement résolus (figure 1.5) :

- ▷ les problèmes d'alignement consistant en l'adaptation des perspectives de l'objet virtuel et de l'environnement réel ;
- ▷ la cohérence photométrique qui consiste en la prise en compte des interactions lumineuses entre l'environnement réel et les objets virtuels ;

⁶ Immersion d'objets réels dans un environnement virtuel

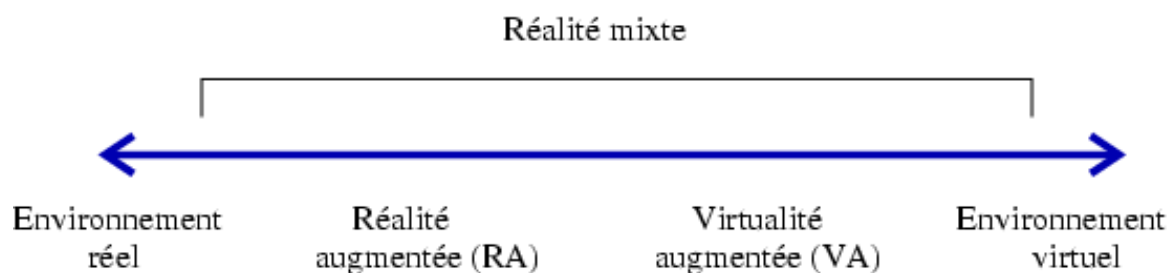


FIG. 1.3 – Le continuum Réel/Virtuel [Milgram 94]



FIG. 1.4 – Vue du système de réalité augmentée ARQuake [Thomas 02] et vue de l'utilisateur.

- ▷ la cohérence spatio-temporelle qui concerne le déplacement des objets virtuels dans le monde réel ainsi que les occultations et interpénétrations qui peuvent se produire ;
- ▷ le cohérence et la crédibilité des comportements. L'insertion d'entités virtuelles possédant un comportement propre implique l'adaptation de ce comportement à l'environnement réel, notamment par rapport aux comportements des entités réelles.

Les environnement réels généralement considérés en réalité augmentée sont statiques : seul l'utilisateur peut s'y déplacer. Nos travaux abordent un problème légèrement différent au travers de l'étude de l'augmentation d'un environnement réel dynamique peuplé d'entités autonomes. Nous nous différencions également par l'augmentation de l'environnement à l'aide d'un acteur virtuel autonome en lieu et places des annotations statiques classiques. Ceci implique des problèmes quant à la gestion de son comportement et son adaptation par rapport aux événements se produisant dans l'environnement.

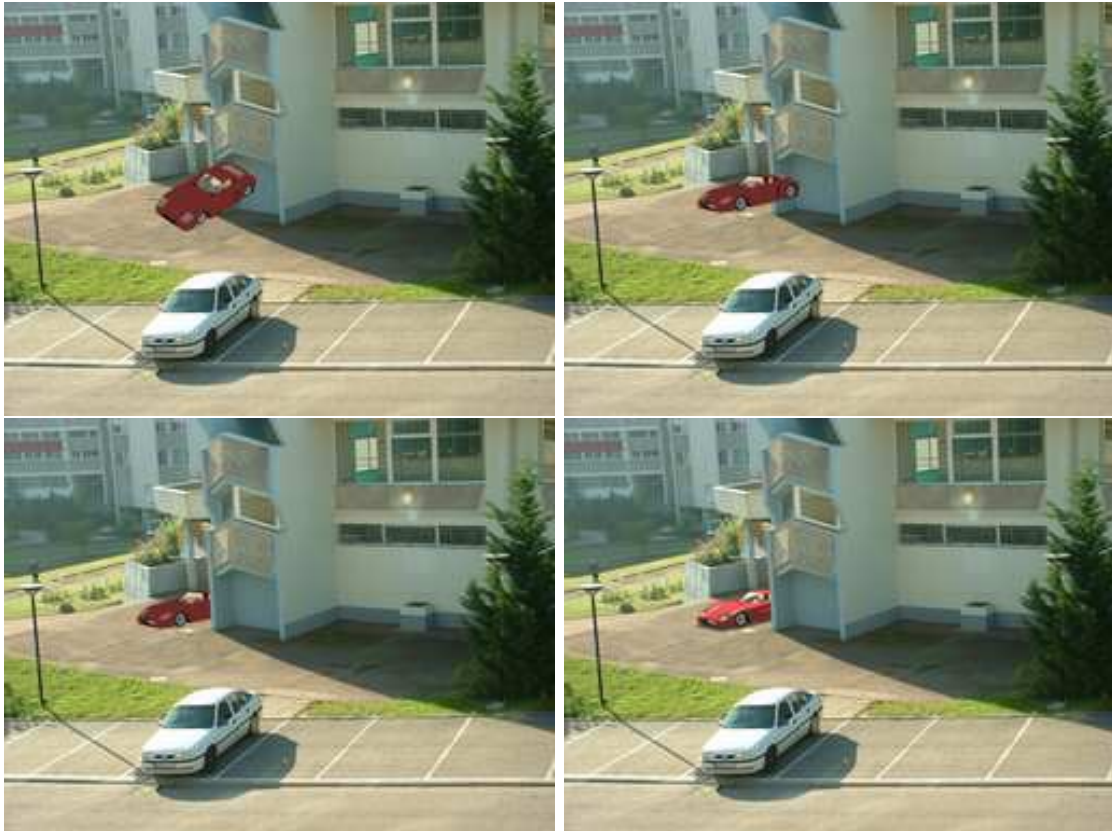


FIG. 1.5 – Illustration des problématiques classiquement associées aux systèmes de réalité augmentée (de haut en bas et de gauche à droite) : incrustation arbitraire, prise en compte des problèmes d'alignement, prise en compte de la cohérence spatio-temporelle et enfin prise en compte de la cohérence photométrique. Images extraites de [http : //www.loria.fr/~gsimon/ra/](http://www.loria.fr/~gsimon/ra/).

1.1.4 Agent autonome et acteur virtuel

La notion d'agent autonome est complexe et sa définition varie suivant les auteurs. Franklin et Graesser [Franklin 96] proposent une classification des agents autonomes suivant deux étapes. La première se base sur une classification générale des agents autonomes au sens large : les agents biologiques, les robots, les agents computationnels. Ces derniers se divisent alors en agents logiciels et vie artificielle. Cette classification représente un point de vue sur les agents. Pour [Nwana 96], définir un agent autonome n'est pas forcément une solution et l'auteur propose une typologie des différentes implémentations d'agents dans laquelle on peut retrouver : les agents collaboratifs, les agents interface, les agents mobiles, les agents d'information, les agents réactifs, les agents hybrides et les agents intelligents. L'auteur lui-même s'accorde à dire que ceci ne constitue encore une fois qu'un point de vue sur la notion d'agent.

D'un point de vue général, la définition proposée par [Franklin 96] nous semble appropriée :

“An autonomous agent is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future.” – [Franklin 96]

On peut distinguer deux grands types de travaux en ce qui concerne les agents [Nwana 96] : les travaux s’intéressant à des problèmes macroscopiques généralement associés à des systèmes multi-agents et les travaux s’intéressant aux problèmes d’ordre microscopique. Le première catégorie se concentre sur des problèmes tels que la communication ou la mise en place de collaborations entre agents alors que la seconde s’intéresse aux architectures internes des agents. Nous nous situons ici dans la seconde catégorie.

Plus particulièrement, nous nous intéressons à la notion d’*acteur virtuel autonome* [Tisseau 01] : un agent logiciel autonome incarné et situé dans un environnement possédant un comportement propre. Un tel agent reste caractérisé par les capacités suivantes :

- ▷ Percevoir son environnement
- ▷ Prendre des décisions en fonction de ses perceptions et de buts propres
- ▷ Agir dans son environnement

1.2 Problématique et proposition

L’objectif de nos travaux consiste en l’étude de la description d’un environnement réel par un *acteur virtuel autonome*. Ce *guide virtuel* est immergé dans l’environnement qu’il décrit à l’aide de techniques de fusion d’images propres au domaine de la *réalité augmentée*. Nous nous plaçons plus spécifiquement dans le cadre des musées et de la structuration d’une description sous la forme d’une visite cohérente. Un tel objectif est ambitieux, il touche directement ou indirectement à une grande variété de problématiques, notamment :

- ▷ la réalité augmentée et les problèmes qui y sont classiquement associés : le calibrage ou l’élimination de surfaces cachées par exemple ;
- ▷ l’animation du personnage virtuel représentant le guide dans l’environnement réel ainsi que les problèmes liés à la gestion des différentes modalités verbales ou non verbales servant à présenter des informations au visiteur ;
- ▷ la structuration de la visite, le choix des informations à présenter en tenant compte de l’aspect pédagogique de la visite et du contexte (visiteur, environnement, état interne du guide, *etc*) ;
- ▷ le comportement du guide virtuel, la façon dont celui-ci va prendre ses décisions en fonction de ses différentes perceptions ;
- ▷ la perception du contexte et plus particulièrement de l’environnement réel dans lequel le guide est immergé.

Dans ce manuscrit nous nous intéressons à un type d'environnement particulier : un environnement dynamique peuplé d'entités autonomes dont le comportement n'est que peu (ou pas) prévisible. Dans une telle situation les problématiques que nous venons d'évoquer restent valables. Néanmoins un tel type d'environnement nous amène à considérer plus particulièrement les aspects suivants :

- ▷ la structuration de la visite – Comment le guide virtuel peut-il s'adapter aux changements de l'environnement réel tout en structurant une visite cohérente pour l'utilisateur ?
- ▷ le comportement du guide virtuel – Étant donné l'aspect dynamique de l'environnement réel dans lequel il doit évoluer, comment le guide virtuel peut-il adapter son comportement afin d'agir de façon crédible et cohérente ?
- ▷ la perception de l'environnement – Face à la complexité de son environnement, quels mécanismes peuvent être mis en place afin de permettre au guide de le percevoir de façon correcte et exploitable du point de vue de la visite guidée ?

Ce sont ces questions que nous développons dans la suite de ce manuscrit. Nous abordons le problème en dotant le guide virtuel d'*autonomie* caractérisée par une boucle perception/décision/action. Lors de la phase de perception, l'agent autonome perçoit l'état de son environnement. Partant de cet état et de son propre état interne il va prendre une décision lors de la seconde phase quant aux actions à mettre en œuvre au cours de la dernière phase. Ces actions vont alors venir modifier l'environnement dans lequel le guide évolue ainsi que son état interne. Dans l'architecture que nous proposons, ces trois phases ne s'exécutent pas de façon séquentielle mais en parallèle au sein de trois modules distincts capables d'interagir les uns avec les autres (figure 1.6) :

- ▷ Le module de perception qui construit une représentation de l'environnement.
- ▷ Le module de décision dont le rôle est de déterminer les explications fournies par le guide ainsi que la façon de les exprimer. Pour cela il fait appel à des connaissances *a priori* et à la représentation de l'environnement.
- ▷ Le module de restitution qui se charge de l'animation et du rendu du personnage 3D qui représente le guide virtuel au travers de modalités visuelles et sonores.

Plus précisément, nous proposons :

- ▷ la notion d'interruptibilité qui permet au guide d'adapter son comportement aux événements pouvant intervenir dans l'environnement.
- ▷ un système d'experts autorisant la définition d'un ensemble de points de vue sur le processus de sélection des explications à présenter aux visiteurs.
- ▷ la mise en œuvre de stratégies de perception par un ensemble de comportements spécialisés afin de construire une représentation de l'environnement réel.

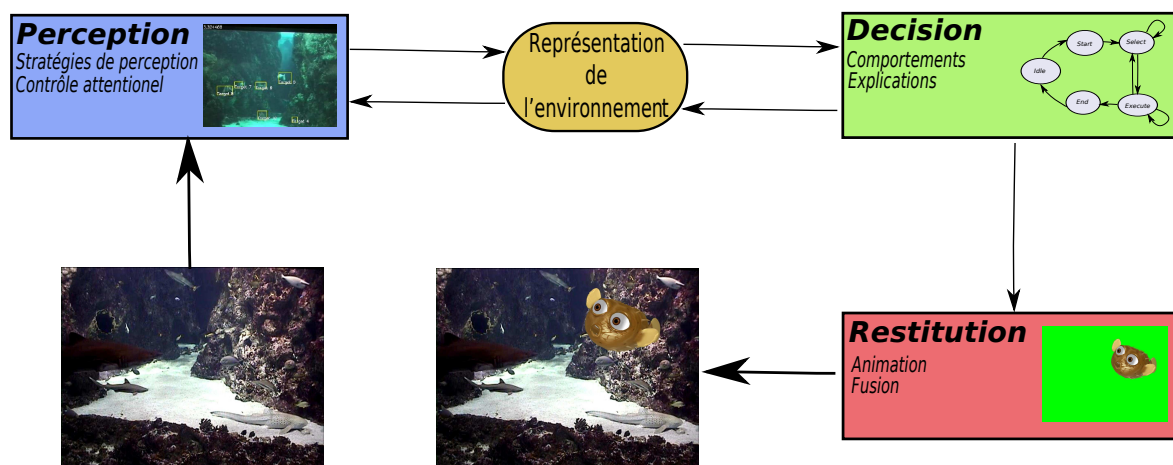


FIG. 1.6 – Vue d'ensemble de notre proposition.

- ▷ un mécanisme de contrôle des stratégies de perception reposant sur la notion d'attention visuelle.

Nous nous intéressons plus particulièrement ici à l'interaction entre le processus de prise de décision et les mécanismes de contrôle de la perception. Nous proposons de mettre en œuvre cette interaction au travers de la représentation de l'environnement et de son interrogation à la manière d'une base de données.

Afin de tester et valider les solutions que nous proposons, nous appliquons notre système à un cas réel : la visite guidée d'un aquarium marin. Un tel cadre présente un excellent exemple d'environnement réel dynamique complexe. Les poissons de différentes espèces qui peuplent cet aquarium sont autonomes par essence et présentent un comportement qui n'est que peu prévisible.

1.3 Organisation du mémoire

La suite de ce mémoire s'organise de la façon suivante :

- ▷ le chapitre 2 présente des travaux liés à la notion de guide virtuel en mettant en avant différents aspects : la visite, l'expression et l'adaptation. Ce chapitre présente ensuite les travaux liés au problème plus particulier de la sélection et la présentation d'informations par un guide incarné dans le cadre d'un environnement dynamique.
- ▷ le chapitre 3 détaille notre approche pour la gestion du comportement et le processus de prise de décision du guide virtuel autonome dans le cadre du choix des explications à fournir au visiteur.
- ▷ le chapitre 4 permet d'appréhender le problème de la perception visuelle ainsi que des travaux concernant la vision par ordinateur. Il détaille ensuite les travaux s'intéressant au cas particulier de la perception active et au problème de l'attention visuelle.

- ▷ le chapitre 5 décrit notre approche pour la gestion de la perception visuelle du guide virtuel autonome ainsi que le mécanisme de contrôle attentionnel qui permet son interaction avec le processus de prise de décision.
- ▷ le chapitre 6 contient les résultats des expérimentations menées dans le cadre de la validation de notre proposition ainsi que des détails en ce qui concerne son application à la visite guidée d'un aquarium marin au sein d'Océanopolis.

Partie II

— Décrire le réel —

Chapitre 2

Du guide réel au guide virtuel autonome

2.1 Introduction

Dans ce chapitre nous présentons différents travaux liés au problème du guide virtuel et de la visite guidée. Pour cela nous commençons par nous intéresser au guide réel et à la façon dont celui-ci met en œuvre une visite guidée pour un groupe de visiteurs. Nous définissons ensuite la notion de guide virtuel au travers de la présentation d'un ensemble de travaux visant à la transmission d'informations à un ou plusieurs utilisateurs.

Considérant la visite guidée comme un problème de construction d'un exposé, nous développons alors les différents aspects de ce problème : le choix des informations à présenter tout d'abord (la structuration de la visite guidée) ; la présentation effective des informations ensuite (l'expression et la communication du guide virtuel).

Enfin nous abordons le problème de l'adaptation d'un point de vue général en nous appuyant sur des travaux liés aux guides virtuels.

2.2 Visite guidée et guide réel

Malgré tous les efforts déployés par les musées afin de fournir des informations aux visiteurs (cf. section 1.1.2), le moyen le plus approprié reste la visite guidée. Une telle visite est par définition limitée :

- ▷ En terme de temps : la visite est préparée pour avoir une durée donnée pendant laquelle un certain nombre d'objets doivent être présentés. Le guide ne peut pas présenter des informations exhaustives sur tous les objets.
- ▷ En terme d'informations : les visiteurs ne sont pas forcément intéressés par l'ensemble des informations disponibles. Un groupe scolaire voudra mettre l'accent sur un point

particulier au long de la visite alors qu’une famille sera plus intéressée par des anecdotes et l’aspect ludique.

Dans le cas qui nous intéresse l’aspect dynamique des objets à présenter est une contrainte supplémentaire. La visite doit être adaptée à ce qui est visible à un instant donné.

L’acteur principal de la visite est bien évidemment le guide. Celui-ci est au cœur de la notion de *docent* que nous évoquions dans la section 1.1. Le dictionnaire de la langue française définit le terme *guide* de la façon suivante :

« *Personne qui accompagne pour montrer le chemin ou donner des explications sur quelque chose.* » – Dictionnaire de la langue française (Bordas) p. 1466

Cette définition met en avant certains aspects importants d’un guide réel :

- ▷ « *Personne* » – Le guide est une personne, il est donc incarné, possède un corps situé dans un environnement. Cela paraît être une évidence lorsqu’on parle des guides humains, ça l’est beaucoup moins lorsqu’on s’intéresse à la notion de guide virtuel (cf. 2.3.1).
- ▷ « *accompagne [...] ou donne[r] des explications* » – Le guide communique des informations aux visiteurs. Ces informations sont adaptées et ont un but pédagogique (on tente de faire comprendre des choses, d’enrichir ses connaissances) sans pour autant tomber dans une approche scolaire : le guide *propose* des informations et ne les impose pas au visiteur.
- ▷ « *explications sur quelque chose* » – Les informations fournies par le guide ont un objet, un support. Le contenu de la visite est ancré dans la réalité du musée, dans l’exposition et dans l’instant de la visite. Ceci est d’autant plus important lorsque la visite concerne des objets dynamiques.

Au cours de la visite, le rôle du guide est de sélectionner et présenter des informations en tenant compte des contraintes liées aux visiteurs et à l’environnement qu’il fait visiter. Pour cela, le guide dispose :

- ▷ d’une « culture » (de connaissances *a priori*) sur les objets de la visite (objets d’art d’une exposition ou bassins et espèces d’un aquarium marin). Cette culture peut être personnelle ou acquise au cours de sa formation.
- ▷ de connaissances sur la visite elle-même : durée, enchaînements, thèmes à aborder en certains point de l’exposition (voir tableau 2.1 dans le cas d’Océanopolis).
- ▷ d’une expérience sur les visiteurs et l’aspect social de la visite guidée : comment gérer un groupe d’enfants agités, comment faire face à un visiteur remettant en cause la place de « professeur » du guide au cours de la visite ou encore comment déterminer l’intérêt principal de tel ou tel groupe de visiteurs.

Lorsqu’il présente des informations aux visiteurs le guide tire parti de ses connaissances et de son expérience afin d’adapter le contenu et la forme de la visite. Il va par exemple faire référence

LIEU	INFOS PARTICULIÈRES	TEMPS	THEME DEVELOPPE
Maquette Requin Baleine	Possibilité d'observer les requins à travers les 3 hublots Se déplacer devant la maquette. Possibilité de la toucher	4 min	Dans cette première partie, on va parler des requins et des raies. <i>Connaissez-vous une différence entre un requin et un autre poisson ?</i> Les requins sont des poissons au squelette cartilagineux et non pas osseux. Leur squelette n'est pas calcifié, il n'est pas dur. [...] Les requins ont en général 2 nageoires sur le dos et 2 nageoires pectorales derrière la tête. Celles-ci servent aux changements de direction. Les requins avancent par ondulation du corps et se propulsent grâce à leur queue. Tous les requins ne se ressemblent pas. On va le voir. La taille, la couleur, la forme changent d'une espèce à l'autre (environ 400 espèces).

TAB. 2.1 – Extrait des documents utilisés à Océanopolis pour diriger la visite guidée du pavillon tropical. La visite totale est prévue pour durer environ 90 minutes.

à une exposition précédente, une réaction du groupe face à tel ou tel sujet ou encore aux visites précédentes afin de mieux faire passer l'information. Le guide tire également parti de l'aspect social de la visite de groupe afin par exemple d'initier des interactions entre visiteurs.

La quantité d'informations présentées est variable d'une visite à l'autre et c'est principalement par expérience que le guide sait quand il doit/peut passer à l'objet suivant (niveau d'attention des visiteurs, temps passé sur un même site, temps restant avant la fin de la visite). La façon dont les informations sont présentées varie également d'un guide à l'autre (chaque guide peut d'ailleurs être « meilleur » pour un « type » particulier de visiteurs). Un aspect plus discutable enfin est celui du point de vue : consciemment ou non le guide transforme les informations qu'il possède au moment de les présenter et y apporte une part de subjectivité.

De ces différents aspects du guide et de la visite, nous retenons les points suivants :

- ▷ La visite – Le guide réel construit une visite de façon dynamique en sélectionnant des informations particulières tout au long du déroulement de celle-ci. Il se base sur ses connaissances (sur les expositions, la visite, *etc*) et son expérience (types de visiteurs, aspect groupe, *etc*) afin de construire une explication cohérente pour transmettre des informations précises.
- ▷ La communication – Le guide réel présente des informations aux visiteurs, pour cela il fait appel aux différentes modalités verbales ou non verbales qui sont à sa disposition. Le guide réel possède un corps qui le contraint dans son environnement mais qu'il utilise pour présenter des informations aux visiteurs.
- ▷ L'adaptation – Tout au long de la visite le guide réel est en adaptation permanente par rapport à une variété de facteurs. L'environnement à décrire (les objets de l'exposition), l'audience (le discours ne sera pas le même suivant le niveau de connaissances du visiteur,

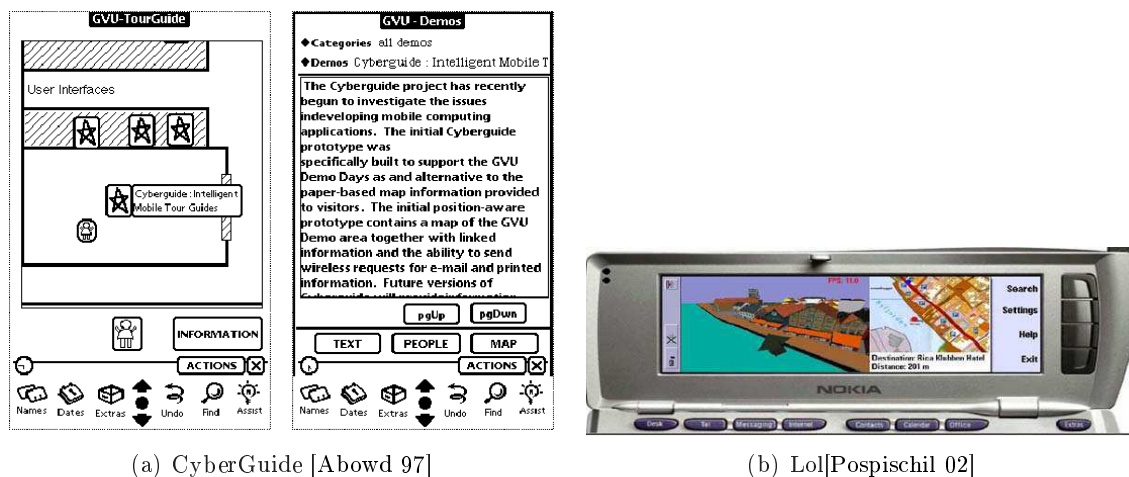


FIG. 2.1 – Exemples de guides mobiles

le nombre de visiteurs ou leur intérêt) et lui-même (l'aspect subjectif du discours, la « touche personnelle du guide »).

2.3 Visite guidée et guide virtuel

2.3.1 Notion de guide virtuel

Nous considérons ici la notion de guide virtuel d'une façon générale au sens d'un système informatique dont l'objectif est d'accompagner et de fournir des informations à un ou plusieurs utilisateurs. Les informations présentées peuvent être diverses : une aide à la navigation dans un environnement réel ou virtuel, des explications précises sur un objet au sein d'un musée, des commentaires sur un environnement dynamique ou encore des informations sur l'utilisation d'un logiciel ou la façon d'effectuer certaines actions. Cette notion large du guide virtuel peut être abordée suivant plusieurs points de vue et domaines : l'informatique mobile, la robotique, les environnements virtuels, la réalité augmentée, l'éducation, la culture ou la formation, les interfaces Homme/Machine au travers de la notion d'agent interface et des agents conversationnels. Ici, nous cherchons à préciser cette notion large du guide virtuel en présentant des exemples dans ces différents domaines.

Le domaine de l'informatique mobile est sans doute celui qui a vu le plus de travaux dans le domaine des guides (figure 2.1. Les possibilités offertes par des systèmes portables comme des PDA et les technologies liées à l'informatique ubiquitaire [Weiser 91] (communication sans-fil entre autres) ont suscité un fort intérêt. Ainsi on trouve des applications permettant d'accompagner le visiteur au sein d'un musée [Bellotti 02; Gool 99; Chou 04] ou d'une exposition quelconque [Abowd 97]. On trouve également des travaux qui s'intéressent à l'aspect touristique en proposant des systèmes capables de fournir des informations pratiques contextualisées par rapport aux choix d'un utilisateur [Malaka 00] pour l'aider par exemple



FIG. 2.2 – Exemples de robots guides utilisés en condition réelles au sein d'un musée [Graf 02].

à planifier sa visite d'une ville [Amendola 04; Cheverst 00] ou l'accompagner pendant son parcours [Pospischil 02]. Les guides mobiles se concentrent sur deux types de problèmes : les architectures matérielles (réseau et communication, périphériques, *etc*) et la contextualisation des informations (par rapport à la position de l'utilisateur, ses préférences, l'historique de ses interactions avec le système, *etc*).

La robotique s'est également intéressée à la notion de guide virtuel (figure 2.2) à travers diverses applications dans des musées [Burgard 99] et des lieux publics [Willeke 01; Nourbakhsh 99]. Ces travaux se sont généralement intéressés en priorité aux problèmes classiques de la robotique comme la navigation dans un environnement comportant de nombreux obstacles statiques et dynamiques [Burgard 99; Thrun 99]. Des travaux plus récents se sont orientés vers l'utilisation de robots en tant que guides en mettant l'accent sur l'interaction multimodale avec le visiteur [Jensen 02; Kim 04a] et l'aspect social de cette interaction [Kuno 07]. D'autres se sont intéressés au ressenti de l'utilisateur grâce à la mise en place de plusieurs robots au sein d'un même musée, chaque robot possédant une apparence et un rôle différent [Graf 02].

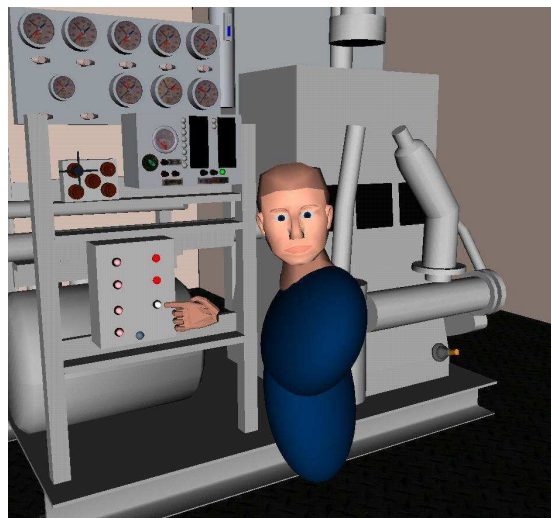
Dans les environnements virtuels (figure 2.3), on retrouve la notion de guide pour l'accompagnement [Doyle 97a; Wachsmuth 97] ou la navigation [Popovici 03]. Ici on s'intéresse à la façon d'associer des informations à l'environnement, sous forme d'annotations par exemple. Dans d'autres travaux, c'est la visite elle-même et la présentation des informations qui sont au cœur des préoccupations [Yuan 03; Chittaro 03].

Les technologies liées à la réalité augmentée ont également attiré l'attention de travaux portant sur la transmission d'informations contextualisées à l'utilisateur¹ (figure 2.4). Soit par simple ajout d'informations statiques (annotations textuelles par exemple) sur un environnement réel en intérieur [Paternier 07; Nagao 95; Schmalstieg 05] ou en extérieur [Hollerer 99a].

¹ Un exemple classique d'applications de la réalité augmentée à la présentation d'informations contextualisées est celui des casques de pilotes de chasse.



(a) Elva au sein du musée virtuel qu'elle fait découvrir aux utilisateurs [Yuan 03].



(b) L'agent pédagogique STEVE qui indique un objet à l'utilisateur [Rickel 98]

FIG. 2.3 – Exemples de guides virtuels en environnement virtuel

Soit en proposant des informations multimédia [Schiele 01] ou même en intégrant des personnages virtuels dans le monde réel [Braun 03]. Notons ici que ces personnages peuvent ou non permettre d'incarner le guide virtuel, ceux-ci servant parfois simplement à « peupler » le monde réel sans réellement s'adresser à l'utilisateur. Le projet ArcheoGuide [Vlahakis 02] quant à lui fait appel à la réalité augmentée pour faire visiter le site archéologique d'Olympie.

Dans le cadre de l'éducation et de la formation, la notion de guide virtuel se retrouve au travers des agents pédagogiques [Rickel 98; Stone 96; Johnson 00]. Ici on cherche à faire acquérir des connaissances à l'utilisateur plus que simplement lui proposer des informations. Dans le domaine culturel, on trouve des applications au sein d'un musée reconstitué en 3D [Yuan 03; Chittaro 03], un lieu historique [Doyle 99; Braun 03], d'un site archéologique [Pollefeys 01] ou de sites virtuels imaginaires [Ibanez 03; Kadobayashi 98].

On peut également aborder la notion de guide virtuel du point de vue de l'interface et de la présentation d'informations [Andre 98]. Au travers d'applications comme des musées virtuels sur Internet [Marucci 00b] ou encore la présentation de lieux tels qu'une université [Panayiotopoulos 99] ou un laboratoire [Björk 98]. Sur le web on retrouve également des travaux dans le domaine culturel avec notamment la présence de guides pour des musées virtuels [Doyle 99; de Almeida 03; Moraes 99]. Ici les problèmes traités portent le plus souvent sur la communication et la gestion des différentes modalités disponibles. La métaphore du guide a également été utilisée afin d'aider un utilisateur à naviguer au sein d'un espace d'information complexe (base de données par exemple) [Salomon 89; Don 91; van Riel 06].



FIG. 2.4 – Exemple de guide virtuel faisant appel à la réalité augmentée en extérieur [Braun 03]

2.3.2 La visite : la construction d'un exposé

Si l'on considère le guide comme un système capable de transmettre des informations à un ou plusieurs utilisateurs, on peut considérer la visite guidée comme la succession de ces informations au cours du temps. Construire cette visite consiste alors en la sélection et la présentation de ces informations.

Ce problème du choix et de la présentation d'informations dans le but de transmettre des connaissances à un ou plusieurs utilisateurs a été abordé dans le cadre des Environnements Informatiques pour l'Apprentissage Humain (EIAH) [Tchounikine 02]. Ces environnements s'intéressent à la mise en place de processus didactiques au sein d'un système informatique. Dans le cas des environnements virtuels ce domaine rejoint souvent celui des agents conversationnels animés (ACA) au travers de la notion d'agent pédagogique [Pelachaud 04; Johnson 00]. L'idée des EIAH est de faire appel à des moyens informatiques afin de mettre en place des procédures d'apprentissage selon un objectif pédagogique. Pour cela les travaux de ce domaine s'attachent à l'étude de plusieurs problématiques telles que la modélisation des connaissances (sur le domaine enseigné, l'apprenant, les méthodes de résolution de problèmes) ou encore l'interaction entre le système et l'utilisateur [Tricot 99; Buche 05]. La modélisation des connaissances et du processus d'apprentissage a donné lieu à la notion de scénario pédagogique [Pernin 04] ainsi qu'à plusieurs langages de modélisation de tels scénarios [van Es 06; Koper 04]. Le but de ces systèmes est de faire acquérir à des apprenants des connaissances ou des gestes techniques définis par des objectifs pédagogiques. Le problème de la visite guidée est ici sensiblement différent. Le but du guide virtuel n'est pas de faire acquérir des connaissances au travers de sessions fortement scénarisées et de contrôler cette acquisition mais de transmettre de façon diffuse (ludique et informelle) des connaissances générales. De plus comme nous l'avons précisé en introduction de ce manuscrit nous ne cherchons pas ici à mettre en place des stratégies pédagogiques pour le visiteur ceci constituant un problème à part entière mais nous nous concentrons sur les moyens d'adapter la transmission d'informations à l'aspect dynamique de l'environnement complexe au sein duquel notre guide évolue. Pour cela nous ne considérons

pas la visite guidée d'un point de vue pédagogique mais d'un point de vue discours c'est à dire comme la sélection et la structuration d'informations au cours du temps.

Du point de vue des guides virtuels, il est généralement admis que la structure narrative est celle utilisée lors de la présentation d'explications durant une visite guidée. Ceci provient notamment du fait qu'il est plus facile pour l'utilisateur de retenir les informations présentées pendant une visite si celles-ci sont globalement organisées, *i.e.* si la visite ne consiste pas uniquement en un ensemble d'informations mises bout à bout de façon arbitraire [Doyle 99]. Un domaine intéressant de ce point de vue pour la structuration d'un discours est celui du *storytelling*. Il existe des travaux concernant la formalisation des structures narratives mais cela reste au delà de notre problématique [Louchart 04]. Sur l'aspect dynamique, des travaux concernent l'émergence d'une histoire au travers des interactions entre les différents protagonistes de cette histoire [Aylett 99]. Cette approche d'*emergent narrative* s'attache à l'étude du problème de l'improvisation et de l'interaction avec un utilisateur humain au sein du storytelling. Ce problème est également abordé dans des travaux sur le storytelling interactif tels que [Franz 03; Cavazza 02; Fred 04]. Ici encore ce problème reste un domaine à part entière et ne concerne qu'indirectement notre étude. Le lecteur pourra se référer à [Mateas 98] pour une vue d'ensemble du problème de la narration et de son étude.

Le point du vue du discours a été abordé en informatique au travers de travaux sur la génération de langage naturel [Bateman 03]. Ceux-ci cherchent à générer des textes structurés et cohérents à partir de données sur les informations à transmettre et sur la structure du discours lui-même. Généralement, ces travaux s'intéressent uniquement à des productions sous forme textuelle. Le problème de leur génération automatique étant abordé d'un point de vue linguistique de façon descendante. Partant d'un but communicatif et d'un ensemble de contraintes (style, langue, longueur, ...), le discours est généré complètement en utilisant tout ou partie des informations disponibles. C'est le cas des schémas proposés par McKeown [McKeown 82; McKeown 85] qui s'appuient sur les aspects structurels communs à différents types de discours afin d'en automatiser la génération à partir d'une base de connaissances. La structure rhétorique du discours issue du domaine de l'analyse formelle de textes [Mann 87] a donné lieu à d'autres approches basées sur la planification globale du discours [Hovy ; Hovy 93]. L'approche proposée par [Grosz 86] explicite la notion d'intention en découpant le problème de la génération automatique de discours suivant ses aspects informationnel, attentionnel et intentionnel. Cette proposition est d'ailleurs plus générale que ce à quoi nous nous intéressons ici car elle intègre le problème de la conversation (aspect attentionnel). Les auteurs proposent également une théorie quant à la cohérence locale du discours [Grosz 95]. Il est intéressant de noter que suivant le type de discours envisagé (exposé ou dialogue par exemple), les structures de données utilisées peuvent varier [Redeker 00]. Ces approches que nous qualifions de « classiques » pour la génération de discours et/ou présentation sont avant tout destinées à la création statique² de documents textuels. Dans le cadre des guides virtuels il est courant de présenter des informations à l'aide de modalités multiples.

Ce problème de la multimodalité a donné naissance au domaine des systèmes intelligents de présentation multimédia [André 00b; Colineau 03] qui est en quelque sorte une extension de la génération de langage naturel³. Les travaux effectués dans ce domaine se sont intéressés

² Certains travaux se sont malgré tout intéressés au problème de génération de textes à horizon restreint [Mellish 98].

principalement à la gestion (choix, synchronisation, ...) de modalités et média multiples généralement dans le cadre de la génération de documents ‘« classiques »’ (rapports, ...). Le domaine est néanmoins relativement vaste et bien qu’il trouve ses racines dans la génération de langage naturel, il se concentre sur une problématique plus générale : celle de la présentation d’informations et notamment l’utilisation de personnages animés pour effectuer cette présentation. Les systèmes intelligents de présentation multimédia introduisent ainsi des problématiques supplémentaires par rapport à la génération « classique » de discours : choix et synchronisation des modalités/médias, gestion de comportements (pour les personnages animés), contraintes temps réel et pression temporelle pour la construction de discours.

Un modèle générique de tels systèmes a été proposé dans [Bordegoni 97]. Ce modèle est destiné explicitement à la présentation des informations même s’il n’exclut pas complètement l’interactivité contrairement aux approches classiques de la génération de langage naturel. Ce modèle se veut modulaire et suffisamment abstrait pour être utilisé dans différentes applications. Plusieurs couches se partagent un ensemble de sources d’information pouvant être organisées suivant quatre groupes : application, utilisateur, contexte et *design*. Chacune de ces sources correspond à un aspect particulier de la génération de présentations multimédias. Les systèmes réels (instanciations de ce modèle générique) ne comprennent généralement pas l’ensemble des aspects du modèle et se limitent soit à l’étude d’un aspect particulier (une couche spécifique du modèle générique) soit à l’intégration de systèmes hétérogènes au sein d’une même architecture. Les applications du modèle générique sont d’ailleurs très variées [André 00b] : génération de rapport, documentation technique, indications routières, gestion de projet, éducation, kiosques d’information, *etc.* Il est important de noter que malgré un modèle apparemment très modulaire, les différentes couches qui le composent sont souvent fortement liées entre elles au sein des applications réelles [Colineau 03].

2.3.3 Bilan

Que ce soit pour une restitution à l’aide de modalités multiples ou pour le langage naturel, nous pouvons identifier les grandes étapes suivantes comme constituant le processus de génération :

- ▷ La sélection de contenu – Le problème du choix des informations à présenter. Généralement par rapport à un but communicatif et un thème en se basant sur l’ensemble des contenus possibles/envisageables.
- ▷ La planification – La phase de structuration du discours durant laquelle les informations sélectionnées sont organisées les unes par rapport aux autres au sein d’un discours cohérent en prenant en compte un ensemble de critères (par exemple le style de discours souhaité).
- ▷ La réalisation – Phase durant laquelle le discours en tant que tel est créé. Des phrases correctes sont générées (ou sélectionnées) puis présentées à l’utilisateur suivant les différentes modalités disponibles.

³ Certaines problématiques sont toutefois complètement différentes comme pour la validation de théories linguistiques par exemple qui n’est généralement pas un objectif des systèmes intelligents de présentation multimédia.

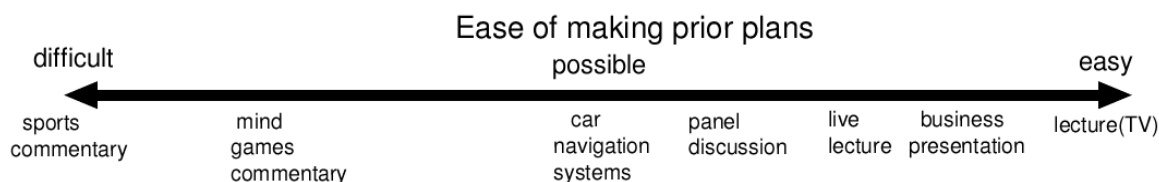


FIG. 2.5 – Facilité de créer des plans *a priori* pour la génération automatique de discours [Tanaka-Ishii 00].

Nous nous intéressons à la génération d'exposés (ou plutôt de présentations multimédias) dans le cadre de contraintes spécifiques : le temps réel (par rapport à l'environnement et à l'utilisateur) et un objet dynamique (l'environnement que l'on souhaite décrire). Comme le montre la figure 2.5 de telles contraintes nous placent dans une situation dans laquelle la définition d'un plan *a priori* est difficile voire irréalisable. Les phases de sélection de contenu et de planification seront inévitablement affectées par cette contrainte. Nous nous plaçons également dans le cadre d'un acteur virtuel autonome, un personnage incarnant le système, ce qui va influencer la présentation d'informations.

2.4 Structuration de la visite

Considérant la visite guidée comme la structuration d'un exposé cohérent, plusieurs approches sont envisageables. Celles-ci dépendent non seulement du domaine d'application mais également de la forme, du support du guide (robot, terminal mobile, personnage dans un environnement virtuel,*etc*), de son environnement (réel ou virtuel, dynamique ou non, *etc*) ainsi que de son objectif (présentation succincte d'informations, interface, éducation,*etc*). Parmi les travaux existants, nous distinguons les types de visite suivants :

- ▷ La visite statique dont la structure est décidée *a priori*, indépendamment de son exécution.
- ▷ La visite interactive dont la structure est déterminée par les interactions explicites entre l'utilisateur et le système.
- ▷ La visite réactive durant laquelle le guide fournit des informations de façon automatique (approche d'un point particulier par exemple).
- ▷ La visite dynamique dont la structure n'est pas définie *a priori*. Ici le guide sélectionne les informations à présenter au fur et à mesure du déroulement de la visite en tenant compte du contexte (utilisateur, environnement, *etc*).

Néanmoins, la plupart des travaux ne se contentent pas d'un type particulier de visite et proposent à l'utilisateur différentes options. On trouve par exemple des systèmes proposant des visites statiques ou interactives [Popovici 03; Pospischil 02]. D'autres mélangent réactivité et interactivité [Hollerer 99a; Abowd 97]. Enfin, certains combinent statisme et réactivité

[Cheverst 00]. Dans [Kadobayashi 98] différents types de visite sont même identifiés. La sélection d'une visite particulière dépend du parcours du visiteur au cours de l'exposition qui précède l'interaction avec le guide virtuel. Nous présentons certains de ces travaux en considérant un des types de visite qu'ils proposent lorsque cela nous paraît pertinent.

2.4.1 Visite statique

La visite statique consiste en une visite dont la structure est prédéterminée, éventuellement personnalisée (cf. section 2.6.1), et ne peut pas être modifiée au cours de son déroulement. Ce type de visite se retrouve dans le cas des robots guides. Dans [Burgard 99] le visiteur se voit présenter une interface proposant différentes visites préprogrammées (plus ou moins longue, incluant différents objets de l'exposition). Dans [Jensen 02] plusieurs robots attendent à l'entrée d'une exposition. Lorsqu'un robot détecte un visiteur, il se présente puis débute une visite entièrement définie *a priori*. De même dans [Nourbakhsh 99] la visite consiste en une succession prédéfinie d'explications d'une durée d'environ trente minutes. Certains robots guides demandent à l'utilisateur de choisir les objets de l'exposition qu'il souhaite voir avant de démarrer la visite [Kim 04a], ces choix sont ensuite respectés sans modification. Sur le web, le système présenté dans [Panayiotopoulos 99] permet au visiteur d'une université virtuelle de déclencher une visite guidée préprogrammée.

Bien que la visite soit planifiée et sa structure prédéterminée, on trouve des exemples dans lesquels son déroulement est adapté, notamment par rapport à l'utilisateur. En réalité virtuelle, [Doyle 99] présente le *Kyoto Tourist Guide* qui permet à un groupe d'utilisateurs de visiter la reconstitution tridimensionnelle d'un château japonais. La visite est découpée en lieux. Le temps passé sur chaque lieu est fonction des réactions des utilisateurs mais chaque visite présente successivement tous les lieux dans un ordre déterminé *a priori*.

2.4.2 Visite interactive

La visite interactive est structurée à partir des actions du visiteur, c'est lui qui choisit explicitement ce qu'il veut voir. C'est le cas notamment pour les systèmes comme les guides touristiques mobiles et les agents conversationnels destinés à l'accueil. Ainsi [Bellotti 02] propose un système mobile dont le fonctionnement s'apparente à celui d'un kiosque d'information : le guide mobile est un système hypermédia portatif et les informations sont présentées au visiteur à sa demande. Dans [Gool 99], le visiteur dispose également d'un appareil portatif équipé cette fois-ci d'une caméra. L'utilisateur pointe la caméra sur l'objet qui l'intéresse, celui-ci est alors identifié par le système qui propose des informations détaillées sur l'objet en question. De même, le système Ubiquitous Talker [Nagao 95] fait appel à la réalité augmentée pour présenter au visiteur des informations sur une partie spécifique d'un objet exposé. Le visiteur sélectionne cette partie en la plaçant dans le champ de vision d'une caméra et les informations sont alors ajoutées à la réalité sous forme de texte sur un écran déporté ; un commentaire audio permet de compléter la présentation.

Parfois le guide peut proposer des informations de sa propre initiative, le déroulement de

la visite et le choix effectif des informations restant malgré tout entre les mains de l'utilisateur. C'est le cas par exemple du système *Web Watcher* proposé dans [Joachims 97]. Les pages web visitées par l'utilisateur sont augmentées par des informations provenant du guide (mise en valeur de liens hypertextes, ajout d'une liste de nouveau liens, *etc*) mais c'est l'utilisateur qui décide de la façon dont il effectue sa navigation et s'il veut ou non consulter les informations proposées. Dans les environnements virtuels, le guide virtuel Merlyn [Doyle 97a] est capable grâce à un environnement annoté de proposer des informations à l'utilisateur mais lui laisse le choix de les consulter ou non. Dans [Doyle 99], les auteurs utilisent le guide virtuel Merlyn pour accompagner des enfants lors de la visite d'un musée web. Ici le guide propose des informations sous la forme d'une histoire mais l'utilisateur garde le contrôle sur le déroulement de cette histoire et sur les informations qui lui sont fournies. Il peut ainsi décider de passer ou non à l'étape suivante de la visite ou demander des informations de façon explicite sur une œuvre particulière. Dans [Marucci 00c] le guide est présent sur un site web sous la forme d'informations supplémentaires présentées séparément et c'est l'utilisateur qui explore de lui-même le musée virtuel. Dans [de Almeida 03], un personnage animé présente des informations contextualisées au visiteur d'un musée web. Le musée, qui présente un ancien navire portugais, est découpé en plusieurs « pièces » et le guide présente des informations d'ordre général à l'entrée de chaque pièce puis des informations spécifiques sont présentées lorsque le visiteur explore de lui-même les différents objets présents dans une pièce.

Dans [Björk 98], un *chatbot* est capable de dialoguer avec l'utilisateur et de répondre à ses questions à propos du laboratoire pour lequel le guide fait office d'hôtesse d'accueil. Dans [Kopp 05], un agent conversationnel (MAX) est installé au sein d'un musée sur le thème de l'informatique et fournit des informations aux visiteurs en dialoguant avec eux. Ces informations peuvent concerner le musée, les expositions qui sont présentées ou d'autres sujet sans rapport avec le musée. Dans de tels systèmes, nous pouvons considérer comme visite l'ensemble de l'interaction entre l'utilisateur et le système comme par exemple dans le système UbiquiTo [Amendola 04] qui permet à l'utilisateur de demander des informations touristiques telles que les restaurant ou hôtels proches.

2.4.3 Visite réactive

La visite réactive concerne des systèmes qui fournissent des informations aux visiteurs de façon automatique. Ici la structure de la visite n'est pas prévue à l'avance mais contrairement aux approches interactives, les informations présentées ne résultent pas d'actions explicites de l'utilisateur. La présentation d'informations se fait en réaction à l'état de l'environnement (position de l'utilisateur, occurrence d'événements, *etc*). Ainsi dans [Pape 98], des guides virtuels sont placés à certains endroits stratégiques d'un environnement virtuel et fournissent des informations à l'utilisateur lorsque celui-ci est suffisamment proche. De même dans [Popovici 03], un guide humain a la possibilité de disposer des guides virtuels en différents endroits de l'environnement. Ces guides fournissent alors des explications de façon automatique au visiteur lorsque celui-ci reste suffisamment longtemps devant un objet. Dans le domaine de la réalité augmentée, [Schiele 01] présente le *Rememberance Agent*, un système de réalité augmentée capable d'associer des informations au contexte afin de les rejouer automatiquement lorsque l'utilisateur se retrouve plus tard dans une situation identique (face à une œuvre spécifique par exemple). L'association des informations au contexte se fait de façon manuelle (soit par

l'utilisateur lui-même soit par un manipulateur à l'avance) grâce à la définition de points remarquables de l'environnement (une peinture par exemple) pouvant être retrouvés ultérieurement à l'aide d'un système de reconnaissance d'image. En réalité augmentée, le système MARS [Hollerer 99a] permet à l'utilisateur de visualiser des informations associées à l'environnement. La position et l'orientation de l'utilisateur sont détectés à l'aide d'un système GPS. Dans [Schmalstieg 05] les informations sont superposées automatiquement au monde réel grâce à un PDA possédant un système de positionnement local à l'intérieur d'un bâtiment. Ces informations consistent en des indications permettant d'aider l'utilisateur à se déplacer dans ce bâtiment.

Même si le déclenchement de la présentation des informations se fait de façon automatique, le visiteur peut néanmoins conserver une liberté dans la consultation d'informations plus détaillées. Ici la notion de visite réactive se mêle à celle de visite interactive que nous avons présentée précédemment. Dans [Chou 04] des informations sommaires sont présentées automatiquement au visiteur sur un terminal mobile en fonction de sa position dans le musée. C'est ensuite au visiteur de demander explicitement des détails. Dans [Augello 07] et [Stock 04] le système proposé détecte automatiquement l'œuvre qui se trouve face à l'utilisateur et propose alors de consulter des informations sur cette œuvre particulière. Dans le domaine de l'informatique mobile, [Abowd 97] propose *CyberGuide*. Ici, le système présente au visiteur une carte lui indiquant sa position courante ainsi que les démonstrations à proximité. Il peut alors sélectionner une démonstration particulière afin d'obtenir sur celle-ci des informations détaillées.

2.4.4 Visite dynamique

La visite dynamique se caractérise par sa capacité d'adaptation par rapport au contexte (environnement et/ou utilisateur). Ici nous ne parlons pas de l'adaptation de la visite au sens des informations présentées, mais bien de la création dynamique de la visite dont la structure et éventuellement le déroulement ne sont pas fixés à l'avance. Nous distinguons cette notion de visite dynamique des visites de type interactif et réactif du fait que l'adaptation dans la visite dynamique est à l'initiative du guide virtuel et non pas une simple réaction aux choix du visiteur : c'est le système qui choisit ce qui doit être présenté au visiteur.

Dans [Ibanez 04b] le guide choisit de façon autonome quel endroit d'un environnement virtuel présenter au visiteur et comment le présenter en fonction d'un ensemble de facteurs tels que la distance entre le guide et l'endroit en question ou son rôle dans l'environnement virtuel (subjectivité du guide). Dans [Yuan 03] le guide virtuel ELVA fournit une visite dynamique pouvant s'effectuer à l'initiative du guide ou du visiteur. L'utilisateur du système peut choisir de suivre le guide ou de se déplacer de façon autonome dans l'environnement, auquel cas le guide l'accompagne pour lui présenter des informations sur les œuvres de son choix. Un mécanisme de raisonnement à partir de cas permet de déterminer le but du guide virtuel en fonction des actions du visiteur. Ce but est utilisé afin de choisir un plan au sein d'une base de connaissances contenant à la fois des plans de visite complète (liste d'objets) et des plans décrivant le discours (*schemas*, cf. figure 2.6) que doit adopter le guide dans une situation particulière. La réévaluation du plan s'effectue à une fréquence dépendante de l'état du dialogue avec le visiteur. Cette état est déterminé par le gestionnaire de dialogue qui

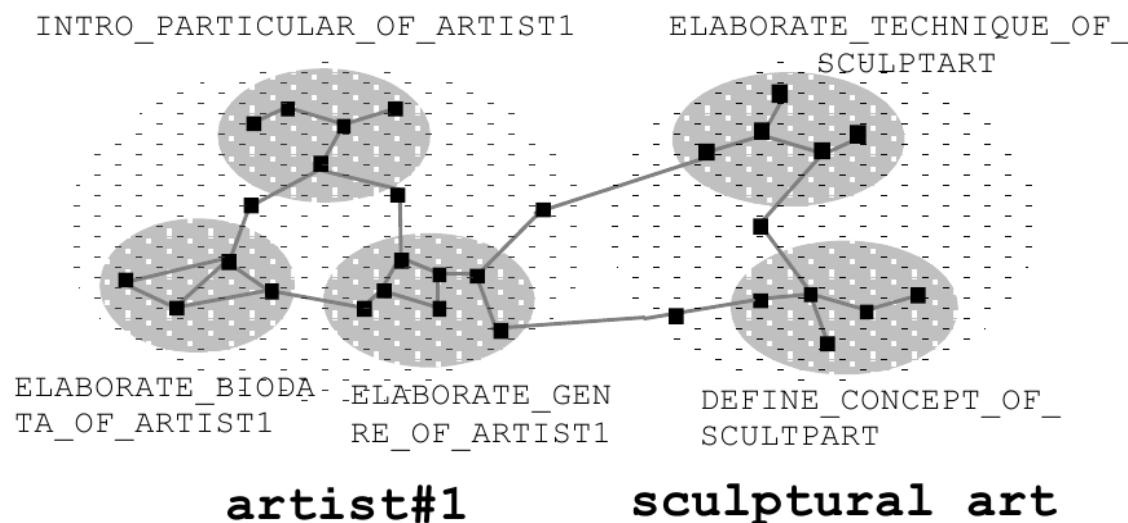


FIG. 2.6 – Vue d'ensemble du discours organisé sous forme de schéma [Yuan 03].

décide de la situation dans laquelle se trouve le guide (début d'une explication, discussion, changement de sujet, *etc*). La transition d'un état à un autre se fait en fonction du contexte (actions, déplacements, phrases du visiteur).

Dans [André 00a] plusieurs systèmes (ROCCO, Byrne et MIKE) capables de générer des commentaires pour la simulation sportive robotique *RoboCup* sont présentés. Ces systèmes utilisent des données limitées sur l'environnement à décrire (un match de football) telles que la position et l'orientation des joueurs, la position du ballon ou encore le score et des informations sur les phases de jeu. A partir de ces informations partielles sur l'environnement, les systèmes présentés se chargent de sélectionner les informations pertinentes en fonction non seulement de l'environnement mais également de l'audience et des modalités disponibles en sortie (généralement ici la modalité verbale).

Le système ROCCO [André 97] (descendant de SOCCER [André 88]) utilise un système de reconnaissance incrémental dans lequel les événements de jeu sont organisés hiérarchiquement (les mouvements et actions sont aux niveaux les plus bas et des notions telles que l'attaque ou la défense se trouvent aux niveaux supérieurs). A partir de ces événements et de connaissances *a priori*, ROCCO utilise une approche incrémentale pour la génération du discours. Le temps est discrétisé et à chaque pas le système sélectionne l'événement le plus intéressant (en fonction du temps écoulé depuis son occurrence et d'une importance intrinsèque) et présente des informations sur cet événement. Si aucun événement n'est disponible il propose alors des informations d'ordre général choisies de façon aléatoire dans une base de données. Il fait également usage d'un historique du discours afin d'éviter les répétitions et gérer la cohérence globale du discours.

Le système Byrne [Binsted 98] se base sur une file de priorités pour sélectionner les événements sur lesquels le discours doit porter. Chaque événement est accompagné d'une

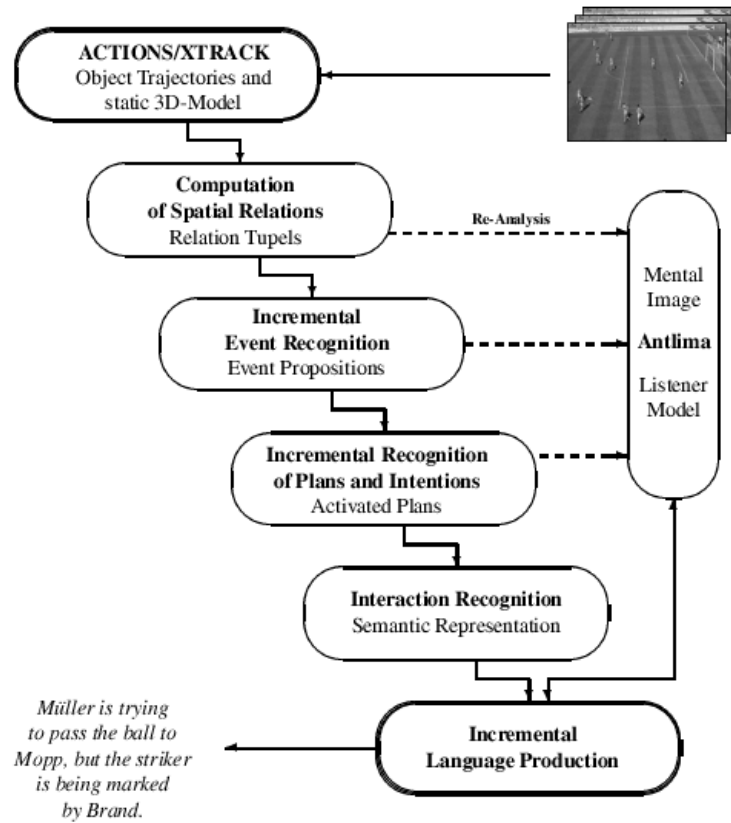


FIG. 2.7 – Architecture globale du système VITRA [Herzog 94].

date de naissance, une date de mort et une priorité. Lorsqu'un événement est sélectionné, un approche basée schéma (modèle de phrase pouvant être contextualisé) est utilisée afin de générer une phrase appropriée par rapport à cet événement (cf. section 2.5.2). Les schémas sont choisis par un système de règles en fonction de la situation du jeu. Si plusieurs sont possibles alors le schéma le moins utilisé jusqu'à présent est sélectionné.

Le système MIKE [Tanaka-Ishii 98] utilise lui un système d'intérêt. Un ensemble d'analyseurs (systèmes de reconnaissance) se chargent de créer des événements qui sont placés dans une mémoire. Cette mémoire est consultée par le commentateur qui possède un ensemble de commentaires possibles sous forme de schémas. A chaque commentaire un score d'importance est associé. Ce score représente intuitivement la quantité d'informations transmises par chaque commentaire. A chaque pas de temps le système instancie des commentaires en fonction du contenu de la mémoire puis le commentateur sélectionne le commentaire apportant le plus d'informations en fonction de son score d'importance. Dans [Tanaka-Ishii 00] les auteurs du système MIKE proposent l'utilisation de plusieurs agents pour la réalisation du discours, chaque agent s'intéressant à un niveau de discours particulier. Ils ajoutent de cette façon la prise en compte du but communicatif en spécialisant chaque commentateur : l'un sur les actions instantanées (passes, buts, *etc*), l'autre sur le jeu à long terme (phases d'attaque ou de défense, *etc*).

Dans le même ordre d’idée on trouve des travaux tels que ceux présentés dans [Herzog 94] qui s’intéressent à la description de scènes réelles par le langage en combinant des systèmes de reconnaissance et de génération de langage naturel. Les auteurs de ces travaux voient en leur système (VITRA pour VIsual TRAnslator) plusieurs types d’applications possibles allant du commentaire sportif (VITRA-SOCCER [André 97]) à la description de scènes de trafic routier. La reconnaissance se fait de façon incrémentale et consiste en l’identification d’événements de différents niveaux (comme pour ROCCO). Ces événements servent de base à la reconnaissance d’interactions qui sont utilisées pour sélectionner les informations à présenter. Les informations générées sont destinées à un utilisateur qui ne voit pas la scène décrite mais on suppose qu’il possède des connaissances *a priori* sur cette scène (type de scène par exemple). Le système exécute un « modèle mental » de l’utilisateur afin de déterminer l’influence des informations présentées sur sa compréhension de la scène décrite. La figure 2.7 présente l’architecture générale du système VITRA.

La structuration dynamique de la visite peut également prendre la forme d’une modification d’un plan de visite existant. En robotique par exemple, RHINO [Burgard 99] propose un mode de fonctionnement dans lequel le robot guide peut être contrôlé par plusieurs utilisateurs sur Internet. Chaque utilisateur peut demander à visiter un point précis du musée, son choix étant ajouté à une file d’attente. Un mécanisme de planification se charge ensuite d’ordonner les requêtes afin de minimiser la distance que le robot va parcourir. Pour le robot MINERVA [Thrun 99] la visite est construite au fur et à mesure de son déroulement en fonction de contraintes temporelles et « pédagogiques ». Les différents points de l’exposition sont organisés sous forme de paires cohérentes en fonction des thèmes abordés en chacun de ces points. Le déroulement de la visite s’effectue de point en point, deux points consécutifs devant appartenir à une même paire cohérente. La visite est construite afin de minimiser les temps de déplacement entre les différents objets présentés. Pour cela, des connaissances temporelles sont récoltées lors d’une phase d’apprentissage impliquant des visites guidées statiques. Cet apprentissage continue ensuite lors des visites réelles.

2.4.5 Bilan

Le rôle d’un guide est d’effectuer une visite guidée afin de fournir des informations aux visiteurs. Au cours de cette visite il construit et structure un discours cohérent en rapport avec l’environnement à décrire. Nous venons ici de nous intéresser à ce problème de la construction d’une visite au travers de la façon dont celle-ci peut être perçue au sein des travaux concernant les guides virtuels tels que nous les avons définis dans la section 2.3.1.

Nous avons identifié et présenté plusieurs types de visite : statique, interactive, réactive, et dynamique. Plusieurs remarques peuvent être faites :

- ▷ la notion de visite interactive permet de rendre le visiteur actif dans le processus de transmission d’informations : c’est lui qui décide ce qu’il veut voir.
- ▷ la visite réactive permet de présenter des informations ponctuelles au visiteur mais interdit généralement de faire appel au contexte ou à l’historique afin d’adapter les informations présentées.

- ▷ la visite dynamique se retrouve sous différentes formes. Elle offre moins de liberté au visiteur mais permet de sélectionner des informations appropriées dans des conditions plus complexes que les autres types de visite.

Décrire un environnement dynamique exclut les approches statiques de la structuration d'une visite guidée. Il n'est en effet pas possible de connaître à l'avance les objets qui pourront être présentés. Les approches réactives nous semblent également insuffisantes du fait de la nature même d'une visite guidée : il est indispensable de pouvoir contrôler sa structure générale. Le problème de la génération d'un exposé structuré et cohérent sous pression temporelle forte reste relativement peu abordé dans les travaux que nous avons présenté ici. Les travaux portants sur la génération automatique de commentaires sur des situations en temps réel semblent les plus proches des problèmes qui nous intéressent. Ils sont différents des systèmes purement réactifs puisqu'ils peuvent tenir compte de l'historique des informations présentées ou d'une façon plus générale du contenu d'une mémoire. Ces systèmes sont néanmoins destinés à la présentation d'informations instantanées. Ils cherchent à décrire immédiatement, à commenter ce qui est visible alors que le rôle du guide au sens des musées est de fournir des informations au visiteur en se basant sur le réel (les objets visibles par exemple) et non pas simplement le paraphraser.

2.5 Expression et communication

Le guide réel fait appel simultanément aux modalités verbales et non verbales qui sont à sa disposition pour appuyer son discours, désigner des objets particuliers ou exprimer un état émotionnel. En informatique, cela correspond au problème de la réalisation de l'exposé : la présentation effective des informations sélectionnées. Celui-ci est particulier et fortement lié aux médias utilisés en sortie, la phase de réalisation ne pouvant être gérée de la même façon pour la génération d'un document texte, multimédia ou une présentation par un personnage animé. Nous nous intéressons dans cette section à la façon de restituer les informations sélectionnées, quelle que soit la méthode de sélection employée. Nous nous concentrons sur les problèmes liés à l'utilisation d'un personnage animé pour incarner le guide virtuel.

2.5.1 Incarnation et personification

Le problème de la gestion de modalités multiples est principalement traité en informatique dans le cadre des systèmes intelligents de présentation multimédia (cf. 2.3.2). Généralement le discours généré par de tel systèmes est destiné à être présenté sous la forme d'un document « classique », parfois avec des contraintes temporelles. Une approche différente a vu le jour dans la fin des années 1990 : l'utilisation de personnages animés crédibles pour la présentation des informations. Ce thème a fait l'objet d'études quant à ses effets sur l'interaction avec l'utilisateur, montrant notamment son influence sur la crédibilité et le ressenti de l'utilisateur lors de l'utilisation du système [Koda 96]. Le constat d'une interaction facilitée par l'incarnation du système de présentation [Reeves 96] a donné lieu dans le domaine de l'intelligence artificielle à de nombreux travaux notamment dans le cadre des

agents conversationnels animés [Cassell 01], des agents interfaces [Maes 94; Maes 95] et plus particulièrement des guides virtuels [Nijholt 03b].

L'utilisation de tels personnages peut se faire à différents niveaux et de différentes façons. Dans le cadre des guides virtuels certains travaux, et notamment ceux liés aux guides mobiles, n'intègrent pas du tout la notion d'incarnation. C'est le cas par exemple pour [Chou 04; Bellotti 02; Gool 99; Malaka 00] et les guides mobiles hypermédias qu'ils proposent. C'est également le cas pour le système CyberGuide [Abowd 97], Archeoguide [Vlahakis 02] ou encore le *Rememberance Agent* de [Schiele 01]. Dans ces exemples, le guide n'est pas matérialisé en tant que tel mais apparaît diffusé sur l'ensemble du système. Lorsque les systèmes intègrent la notion d'incarnation, il est possible de distinguer deux approches :

- ▷ La personnalification : le guide est matérialisé mais ne possède pas de corps physique qu'il peut utiliser lors de la présentation d'informations. C'est le cas par exemple dans [Doyle 97a] : le guide virtuel *Merlyn* est un personnage qui évolue au sein d'un jeu en mode texte. Il n'a pas d'existence physique visible dans cet environnement (le type d'environnement ne le permet pas) mais une présence pour l'utilisateur avec qui il peut communiquer. Dans le même ordre d'idée, on trouve des *chatbots* tels que *Viktoria* [Björk 98] qui repose sur la mise en place d'un dialogue sous forme de texte avec l'utilisateur. Les systèmes présentés dans [Cheverst 00] (GUIDE) et [Marucci 00b] utilisent une image fixe qui représente le guide (figure 2.8). Dans [Pollefeys 01] le guide virtuel est incarné par un visage dont les possibilités en terme de présentation d'informations restent relativement limitées (déplacements dans l'environnement virtuel, expressions faciales et parole essentiellement).
- ▷ L'incarnation : le guide possède un corps situé dans un environnement et il utilise ce corps lors de la présentation d'informations. On retrouve cette notion d'incarnation dans une forme relativement simple au sein de travaux concernant les agents conversationnels tels que [de Almeida 03] ou encore [Doyle 99; Moraes 99]. Ici le guide est incarné au travers d'un personnage de type *Microsoft Agent* et peut faire usage de plusieurs modalités. Les robot guides présentent quant à eux une incarnation forte par définition : ils possèdent un corps physique qui les contraint dans l'environnement réel. Certains font usage de ce corps pour transmettre des informations suivant différentes modalités [Thrun 99; Jensen 02; Nourbakhsh 99; Kim 04a]. Dans les environnements virtuels on trouve des exemples de guides fortement incarnés intégrés dans l'environnement de la visite. C'est le cas par exemple pour [Chittaro 03] ou [Yuan 03]. Ici le guide virtuel est un humanoïde qui utilise des modalités verbales et non verbales pour transmettre des informations aux visiteurs. Il est intégré dans l'environnement de la visite et son corps le contraint dans cet environnement (le guide doit par exemple être capable d'éviter les objets). Dans [Kopp 05] le guide possède un corps représenté par un modèle tridimensionnel qui est incrusté dans l'environnement réel (figure 2.9). Il est capable d'utiliser un ensemble de modalités (verbales ou non) afin de présenter des informations et/ou d'interagir avec l'utilisateur.

Dans le cadre d'une application de réalité augmentée, [Braun 03] présente des travaux intéressants du fait que le guide virtuel est dans l'environnement réel à propos duquel il fournit des informations aux utilisateurs.

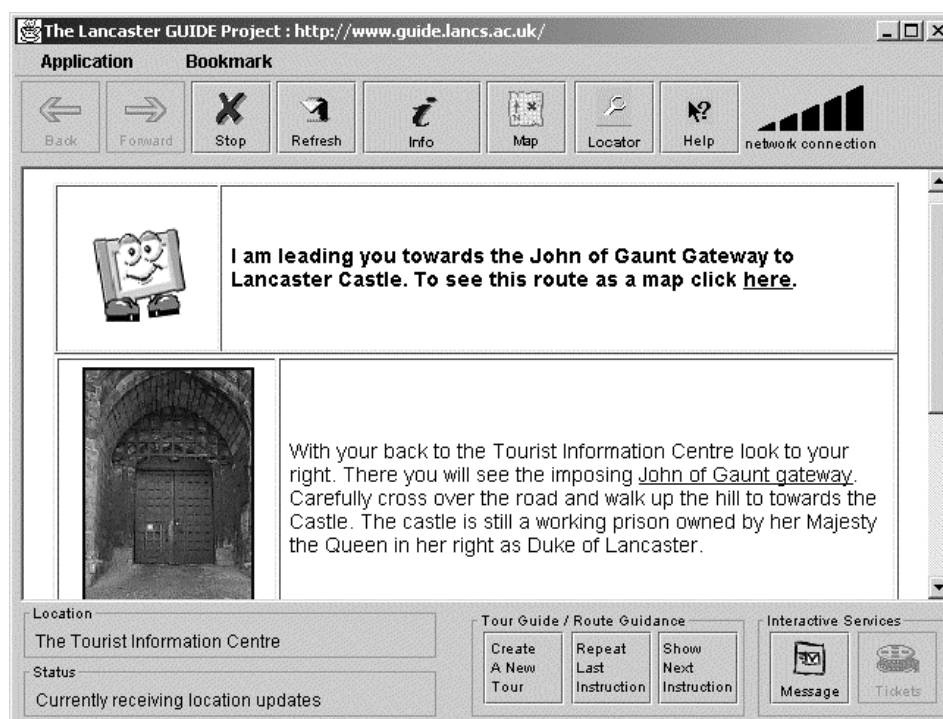


FIG. 2.8 – Capture d'écran de l'interface du système GUIDE [Cheverst 00].

2.5.2 Communication multimodale

L'aspect commun aux différentes utilisations de l'incarnation réside dans les possibilités de communication multimodale. Les personnages incarnant les guides virtuels disposent de capacités de communication spécifiques au delà des possibilités offertes par les média « classiques » (image, son, vidéo). Ces capacités sont celles communément catégorisées comme verbales (la parole) et non verbales (gestes, expressions, positions du corps, *etc*).

Si l'on met de côté la gestion des modalités non verbales et que l'on s'intéresse uniquement à la « parole », on distingue deux approches :

- ▷ Les approches basées grammaire qui font usage de règles linguistiques afin de construire des phrase correctes.
- ▷ Les approches basées modèle qui utilisent des schémas de phrases dont certaines parties peuvent être contextualisées. Ces parties variables étant générées lors de la production du discours.

La communication orale se fait selon deux approches : la restitution de fichiers sons pré-enregistrés ou la synthèse vocale (*Text-To-Speech*) [Lemmetty 99]. La première permet une restitution réaliste au prix d'une flexibilité réduite. La seconde offre plus de souplesse (création des sons à la volée) mais reste limitée en terme de qualité (restitution d'émotions difficile par exemple). A noter que ces limitation font l'objet de plusieurs travaux [Hoult 04; Schröder 01].



FIG. 2.9 – L’incarnation du guide d’accueil présenté dans [Kopp 05].

La communication non verbale repose quant à elle sur des techniques d’animations faciales ou corporelles [Cassell 94; Herwin van Welbergen 06; Beskow 97a]. Ces différents problèmes sont complexes et sont abordés dans des domaines tels que celui des agents conversationnels animés. Ici nous nous concentrons sur l’aspect dynamique de l’environnement et nous ne détaillerons pas plus cet aspect de la communication multimodale. Le lecteur pourra se référer à [Pelachaud 04] pour plus de détails sur ce type de problème.

Au delà des problèmes de réalisation que nous venons d’évoquer s’ajoute celui du choix [Theune 01; Theune 05] et de la synchronisation des différentes modalités [Jonker 99]. Ce problème est également complexe. Il a été notamment abordé dans le cadre des agents conversationnels et des agents de présentation. Nous ne présentons ici qu’un bref aperçu des travaux existants. Dans [Cassell 94], des automates finis parallèles (*PaTNet*) sont utilisés. De façon similaire, [Devillers 02] utilise HPTS [Donikian 01] pour la coordination des gestes d’un humanoïde virtuel. Dans [Noma 00], un langage permet de décrire la synchronisation des différents types de modalités de façon simple. Pour [de Carolis 02] c’est un plan de discours qui est transformé selon des connaissances du domaine en un script d’animation. D’autres travaux se sont attachés au problème de la modélisation des émotions et de la personnalité [Padgham 97].

2.5.3 Incarnation et comportement

L'incarnation du système de présentation à travers un personnage animé crédible implique, en plus de la gestion de différentes modalités, d'intégrer le comportement propre au personnage [Andre 98; André 00b].

Dans [Andre 98] les auteurs présentent deux projets (PPP et AiA) qui visent à la mise en œuvre d'agents de présentation incarnés. Ici le comportement du personnage consiste en la combinaisons de commandes d'animation liées au processus de présentation des informations et de commandes liées à la personnalité du personnage. Ces dernières sont divisées en différentes catégories :

- ▷ les commandes liées à la gestion bas niveau des déplacements du personnage. Dans le cas d'un humanoïde cela peut par exemple consister en la gestion des animations de marche du personnage.
- ▷ les commandes permettant la crédibilité du personnage, notamment lors des phases d'attente durant lesquelles aucune information n'est présentée à l'utilisateur. Cela inclut par exemple le fait de taper du pied ou regarder autour de soi.
- ▷ les commandes devant s'exécuter en réaction à l'utilisateur. Cela peut par exemple consister en un suivi du pointeur de souris de l'utilisateur par le regard du personnage animé.

Les différentes commandes (de présentation et de comportement propre) sont définies sous la forme de primitives et de séquences d'actions. Une primitive est caractérisée par un ensemble de pré- et post-conditions et une séquence est constituée d'un ensemble de pré-conditions et d'une suite de primitives. Ces différentes composantes sont définies de façon déclarative pour permettre ensuite la compilation d'un comportement complet sous la forme d'une machine à états.

Au sein du système *CrossTalk* [Klesen 03] plusieurs agents interagissent avec l'utilisateur au travers du jeu de scènes (scénarios) préprogrammées. Le comportement des personnages est géré à l'aide d'une machine à états, hiérarchique cette fois, qui organise les différentes scènes (figure 2.10). La machine à état est définie à l'aide de deux types de nœuds : les nœuds élémentaires qui correspondent à une scène prédéfinie et les "super-nœuds" qui correspondent à une scène prédéfinie ou une scène générée automatiquement au cours de l'exécution. Un "super-nœud" peut contenir d'autres nœuds sous la forme d'une machine à états locale. Lors de la définition du comportement, les liens d'un "super-nœud" sont automatiquement hérités par l'ensemble des nœuds des niveaux inférieurs. Les liens entre nœuds peuvent être de l'un des types suivants : probabiliste, conditionnel ou interruption. Les liens de type interruption permettent de définir une condition d'interruption du nœud courant autorisant la prise en compte d'événements non prévus (actions utilisateur par exemple) à travers la définition de « raccourcis » au sein du comportement global (liens en pointillés serrés dans la figure 2.10).

Dans [Stone 96] c'est le système pédagogique *DESIGN-A-PLANT* qui est présenté. Un agent pédagogique animé (*Herman the Bug*) y communique des informations sur la botanique. Ici la gestion du comportement repose sur la notion de *behavior space*, un ensemble de

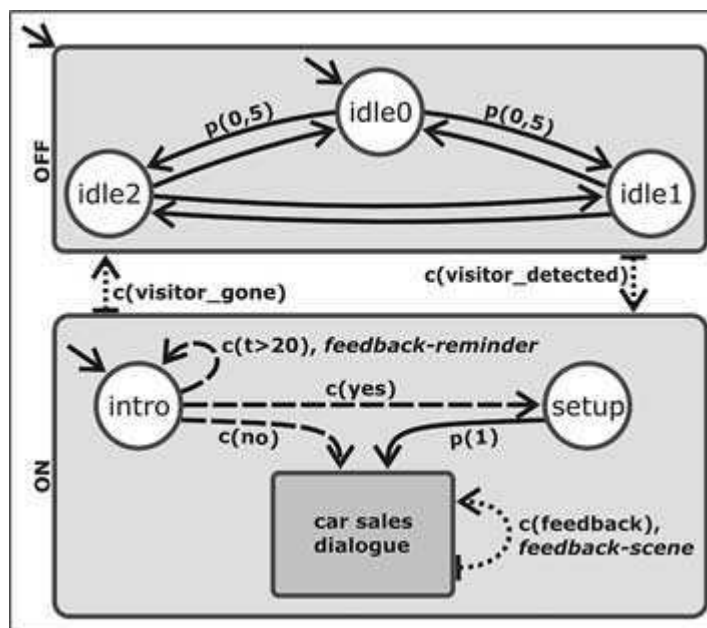


FIG. 2.10 – Comportement sous forme de machine à état interruptible dans [Klesen 03].

comportements simples qui sont combinés au cours de l'exécution en fonction d'un ensemble de contraintes. Chaque comportement consiste en la définition précise des animations et sons devant être joués, les différents comportements étant structurés au sein d'un réseau afin de pouvoir être ordonnés ultérieurement. Cette structuration s'effectue à l'aide d'un index tripartite (ontologique, intentionnel et rhétorique) permettant d'annoter les comportements, la notion de prérequis imposant un ordre partiel sur l'espace des comportements et des informations de continuité visuelle (possibilité d'enchaîner les animations). A partir de l'espace structuré des comportements, le choix du comportement à exécuter à un instant donné se fait en fonction de l'état du système, de l'utilisateur (demande d'assistance, proposition incorrecte, absence d'actions depuis un certain temps, *etc*) et des contraintes de cohérence que nous avons présentées précédemment. Les différents comportements sélectionnés sont ensuite assemblés au sein d'un comportement final qui est exécuté.

2.5.4 Bilan

Nous nous sommes intéressés au problème de l'expression et la communication du guide virtuel. Tout d'abord au travers du problème de l'incarnation, caractéristique indéniable du guide réel. Nous avons noté qu'il est généralement admis que l'utilisation de l'incarnation pour un système destiné à la présentation d'informations ou à l'interaction avec un utilisateur favorise la transmission d'informations en permettant une communication plus naturelle entre l'utilisateur et le système.

L'incarnation du guide implique l'utilisation de modalités multiples. Cette possibilité

apporte un certain nombre de problèmes quant à la réalisation de la présentation des informations (synthèse vocale, animation) ainsi qu'au choix et à la synchronisation des différentes modalités. Ces problèmes sont complexes mais nous semblent néanmoins secondaires dans le cadre de notre étude de la description d'un environnement dynamique. Ainsi, une approche basée sur la notion de script nous semble suffisante et permet de nous concentrer sur le problème de la sélection des informations et de la gestion du comportement.

Incarner le guide au sein d'un personnage animé implique également de devoir gérer son comportement afin qu'il s'intègre de façon crédible dans son environnement. Il ressort des travaux que nous avons présentés la nécessité de pouvoir intégrer le comportement propre du personnage et le comportement de présentation d'informations de façon transparente. Nous avons présenté plusieurs approches pour la gestion du comportement. Parmi celles-ci, l'utilisation de machines à états interruptibles nous semble intéressante dans le cadre d'un environnement dynamique. Il apparaît néanmoins nécessaire de pouvoir contrôler finement ces interruptions, les événements non prévus pouvant intervenir en permanence au sein de l'environnement que l'on cherche à décrire.

2.6 L'adaptation

La visite guidée réelle est en permanence modelée par un ensemble de facteurs (les visiteurs, l'environnement, le guide lui-même). Cette adaptation peut s'effectuer de différentes façons et agir sur différents objets. En informatique ce problème est également abordé au sein de travaux divers.

2.6.1 Personnalisation ou adaptivité

D'une façon générale, on peut distinguer deux types d'adaptation [Fantoni 03] :

- ▷ L'adaptation « hors-ligne » ou personnalisation qui concerne tous les mécanismes mis en œuvre avant l'exécution du système. Cela peut consister en un profil utilisateur ou un mode de fonctionnement spécifique.
- ▷ L'adaptation « en-ligne » ou adaptivité qui concerne les mécanismes mis en œuvre pendant l'exécution du système et lui permettent de s'adapter de façon dynamique aux modifications de l'environnement par exemple. Cela peut consister en une analyse permanente des actions de l'utilisateur ou une surveillance du déroulement de la visite.

La personnalisation considère essentiellement des facteurs relativement statiques (le type ou le niveau d'éducation du visiteur, la personnalité du guide), alors que l'adaptivité est par définition plus adaptée à des facteurs dynamiques (l'environnement ou l'interaction entre le guide et le visiteur au cours de la visite).

2.6.2 Objets de l'adaptation

Quel que soit le type d'adaptation envisagé, celle-ci peut affecter différents aspects d'un système. Dans le cadre de la présentation d'informations nous identifions les objets d'adaptation suivants :

- ▷ L'adaptation de contenu : c'est l'adaptation des informations présentées par rapport au niveau de connaissances ou à l'âge de l'utilisateur par exemple. C'est également l'adaptation de la visite elle-même, du choix des informations à présenter. Cet aspect est généralement lié à l'utilisateur mais peut se voir influencer par d'autres facteurs tels que la durée prévue de la visite ou encore l'état de l'environnement.
- ▷ L'adaptation de présentation : la façon dont les informations sont proposées aux visiteurs est modifiée en fonction du contexte. Un guide va par exemple utiliser des gestes pour présenter tel ou tel point ou désigner un aspect particulier d'une œuvre d'art. Cet aspect est fortement lié au support physique utilisé par le guide. Ainsi dans le cas d'un terminal mobile cette adaptation peut consister en une modification de l'affichage des informations suivant la puissance de calcul du terminal utilisé ou de préférences de l'utilisateur. Dans le cas d'un guide incarné cela peut se traduire par une gestuelle particulière ou la langue dans laquelle les informations sont présentées.

L'adaptation de contenu est présente dans la plupart des travaux existants sur les guides virtuels. Dans [Marucci 00b] par exemple le contenu présenté est adapté en fonction du profil sélectionné par l'utilisateur parmi les choix suivants : touriste, étudiant et expert. Chaque profil détermine un jeu de paramètres pour l'exécution du système. Dans [Marti 99], des commentaires audio sont adaptés en fonction des déplacements des visiteurs au sein du musée, les sons joués correspondant aux objets proches de sa position. L'adaptation de contenu au sens du choix des informations à présenter se retrouve dans le cadre de la visite dynamique que nous avons présentée dans la section 2.4.4.

L'adaptation au niveau de la présentation peut se retrouver dans le cadre d'applications mobiles ou destinées à un ensemble de supports [Amendola 04]. Dans [Vlahakis 02] le système *ArcheoGuide* intègre dès sa conception la possibilité d'interfaces multiples (casque de réalité augmentée, tablette tactile ou PDA). Les fonctionnalités et la restitution sont ensuite adaptées aux possibilités de l'interface utilisée. Dans [Paris 01b; Paris 01a; Lu 00] la présentation d'informations est adaptée à l'appareil utilisé par le visiteur ainsi qu'à ce qui l'intéresse. Parfois un support limité peut même être utilisé pour aider à la personnalisation en forçant l'interaction avec l'utilisateur [Brusilovsky 03]. Ce type d'adaptation peut concerner l'usage de différentes modalités dans le cadre d'un système incarné par un personnage animé (cf. section 2.5.2 et 2.5.3).

2.6.3 Sources d'adaptation

Dans le cas du guide réel, nous avons identifié plusieurs facteurs pouvant influencer la sélection des informations présentées (cf. 2.2) :

- ▷ Le visiteur ou groupe de visiteurs : en adaptant par exemple le niveau des connaissances transmises ou en adaptant la durée ou le contenu de la visite elle-même en fonction de l'intérêt des visiteurs pour un thème particulier.
- ▷ Le contexte extérieur, l'environnement. Dans le cas d'un aquarium par exemple : les poissons qui sont visibles à un moment donné.
- ▷ Le guide lui même : un guide particulier peut avoir un domaine de prédilection ou un avis personnel sur l'exposition qu'il présente.

On peut retrouver ces différents facteurs dans les travaux concernant les guides virtuels.

Beaucoup de travaux sur les guides virtuels se sont attachés au problème particulier de l'adaptation par rapport au visiteur. Au travers de la modélisation de son comportement, de la façon dont il utilise le système mais également de façon plus explicite en demandant à l'utilisateur de fournir au système des informations sur ses préférences et ses centres d'intérêt. Dans [de Almeida 03] par exemple les informations fournies par le guide virtuel sont adaptées en fonction de l'intérêt supposé de l'utilisateur pour des objets particuliers de l'environnement. Cet intérêt est déterminé lors du dialogue entre l'utilisateur et le guide virtuel (questions posées au guide et réponses de l'utilisateur pour le guide). Dans le premier cas, le guide virtuel va pouvoir grâce à sa base de connaissances déterminer les informations pouvant être associées à la question de l'utilisateur et les prendre en compte dans la suite de la visite. Dans le second le guide va poser des questions spécifiques à l'utilisateur telles que : “*Avez vous remarqué cet objet là ?*”. En fonction de la réponse le guide va alors déterminer si l'utilisateur est intéressé ou non par les informations associées à cet objet.

Pour [Doyle 99] ce sont les actions des visiteurs qui déterminent à la fois les informations présentées et le moment auquel le guide va passer d'une étape de la visite à l'étape suivante. Pour cela, le guide attend une “validation” de la part des utilisateurs lui indiquant qu'ils sont prêts à passer au point suivant de la visite. Dans [Thrun 99], le visiteur sélectionne les points particuliers qu'il souhaite voir intégrés à sa visite et le robot guide MINERVA planifie ensuite la visite en fonction de ces préférences et d'autres paramètres tels que la durée moyenne d'une visite calculée à partir des dernières visites effectuées. Les choix des visiteurs sont pris en compte dans cette planification mais peuvent ne pas être totalement respectés en fonction des autres contraintes considérées. Dans [Joachims 97], le système *Web Watcher* permet l'adaptation de pages web au fur et à mesure de la navigation effectuée par l'utilisateur. Le guide va ici ajouter des informations aux pages originales sous différentes formes : liste de suggestions de liens à suivre, mise en valeur de liens en rapport avec les intérêts de l'utilisateur et liste de commandes propres au fonctionnement du guide. L'intérêt d'un lien pour une recherche donnée est déterminé à partir de données obtenues par apprentissage au fur et à mesure des sessions d'utilisation du système. La pertinence d'un lien est obtenue grâce au calcul d'une valeur de similarité entre les mots-clés associés au lien et les mots-clés de la requête utilisateur.

L'adaptation par rapport à l'environnement est particulièrement présente dans les travaux sur les guides mobiles. La position et l'orientation de l'utilisateur sont généralement utilisées afin de contextualiser les informations que celles-ci soient demandées explicitement par l'utilisateur [Abowd 97; Sumi 98] ou présentées de façon automatique [Stock 04; Augello 07]. D'autres facteurs peuvent entrer en ligne de compte comme par exemple les horaires d'ouverture d'un site touristique [Davies 98]. Dans [Raptis 05], les auteurs s'attachent à formaliser

la définition du contexte dans le cadre de ces applications mobiles en général et du musée en particulier. À noter qu'ici les auteurs considèrent plutôt un aspect matériel (les différents terminaux, l'architecture de communication, *etc.*). En réalité augmentée, [Grafe 02] sélectionne les informations à présenter suivant l'objet désigné par l'utilisateur à l'aide d'une caméra vidéo. [Schiele 01] présente un exemple typique d'application capable de s'adapter à l'environnement au travers de son *Remembrance Agent*. Celui-ci enregistre l'environnement autour de l'utilisateur et associe ces informations à des éléments clés de ce même environnement tels que des objets ou des sons. Plus tard ces éléments clés sont utilisés afin de rappeler les enregistrements de l'environnement en mémoire et les présenter au visiteur.

Enfin, certains travaux se sont attachés à adapter les informations et à les personnaliser en fonction du guide lui-même. Ainsi, dans [Ibanez 03; Ibanez 04a; Ibanez 04b] les auteurs présentent un système capable d'adapter son discours à un profil défini en terme d'intérêts ou de rôles. Pour cela les explications, structurées sous la forme d'une histoire, sont organisées au sein d'un réseau de faits. Lorsqu'un fait est sélectionné, des informations similaires sont choisies par extension en prenant en compte les caractéristiques du guide virtuel (sa nationalité par exemple). Ces faits sont ensuite ordonnés et transformés pour former le discours du guide. Cette transformation inclut la génération d'attitudes et d'actions associées au texte des explications permettant de communiquer les émotions du guide et/ou d'appuyer son discours.

2.6.4 Bilan

Nous avons présenté différentes facettes de l'adaptation dans le cadre des guides virtuels. Pour cela nous avons tout d'abord différencié l'adaptation *a priori* (personnalisation) de l'adaptation pendant l'exécution (adaptivité). Ces deux types d'adaptation nous semblent complémentaires. Nous avons remarqué que la notion d'adaptivité tend à être de plus en plus utilisée et offre une plus grande souplesse : on s'adapte au fur et à mesure lorsque cela est nécessaire. Nous nous sommes ensuite intéressés aux objets pouvant subir l'adaptation au sein du système, en distinguant le contenu et la présentation. Enfin, nous avons présenté les différentes sources d'adaptation possibles, à savoir : le visiteur, l'environnement et le guide lui-même.

Il apparaît indispensable dans le cadre de la description d'un environnement dynamique de mettre en œuvre un système adaptatif, les sources de l'adaptation étant de fait dynamiques. Néanmoins, l'étude des différentes sources d'adaptation possibles (le visiteur/utilisateur, l'environnement et le guide lui-même) a mis en avant la difficulté de percevoir ces différentes sources et notamment l'environnement et l'utilisateur. Dans le cas de l'utilisateur, cette perception implique l'interprétation des interactions avec le système. Dans le cas de l'environnement et notamment dans le cas d'un environnement réel, cela implique généralement l'utilisation de systèmes de positionnement et/ou de systèmes de vision par ordinateur permettant d'identifier la situation dans laquelle se trouve l'utilisateur.

Dans le cadre de notre étude il est inévitable de s'adapter à l'évolution de l'environnement. L'adaptation par rapport à l'utilisateur nous paraît au contraire secondaire de ce point de vue. Il semble néanmoins important d'être en mesure de proposer un système permettant l'intégration de multiples facteurs d'adaptation.

2.7 Conclusion

Ce chapitre a été l'occasion pour nous de présenter différents aspects de la description d'un environnement dynamique par un acteur virtuel autonome. D'une façon plus générale, nous avons commencé par décrire le guide réel et la façon dont celui-ci effectue la visite guidée, mettant en évidence ses capacités d'adaptation et de transmission d'informations. Nous avons ensuite mis en avant la complexité de la notion de guide virtuel et l'absence de définition claire dans le domaine informatique à l'heure actuelle. Partant d'une définition relativement simple et générale, nous nous sommes alors intéressés aux travaux permettant l'accompagnement ou la transmission d'informations à un (des) utilisateur(s) humain(s). Cette notion large nous a amenés à considérer des travaux allant de la robotique aux environnements virtuels en passant par la réalité augmentée et l'informatique mobile.

Parmi ces divers travaux nous nous sommes alors intéressés à la façon dont les informations devant être présentées sont sélectionnées, comment la visite est structurée. Ceci nous a amenés à considérer différentes approches, parmi lesquelles nous avons noté la pertinence des approches dynamiques. Étant donné le contexte de notre étude, à savoir la description d'un environnement réel dynamique complexe, il nous semble évident qu'une telle approche s'impose. En effet, la visite guidée que nous cherchons à mettre en œuvre doit pouvoir être modulée en fonction des changements de l'environnement. Sa construction doit néanmoins tenir compte des contraintes inhérentes à la structuration d'un discours explicatif pour le visiteur et notamment des contraintes de cohérence. Pour cela il nous semble intéressant de mêler des approches scénarisées aux mécanismes existants dans le cadre de systèmes temps réel.

Nous nous sommes ensuite concentrés sur le problème de la présentation des informations sélectionnées et notamment au problème de l'incarnation, caractéristique importante du guide réel. Nous avons différencié la personnification de l'incarnation en mettant l'accent sur une notion d'incarnation forte au sens où le guide est immergé dans l'environnement qu'il décrit. Les différents problèmes liés à la communication multimodale que nous avons évoqués ont montré la complexité de cette question que nous considérons ici comme au delà de notre problématique. Ces problèmes d'animation et de synchronisation de différentes modalités, bien qu'importants, ne nous semblent pas ici essentiels du point de vue de l'aspect dynamique auquel nous nous intéressons. Il reste néanmoins important de ne pas perdre de vue ces problèmes au travers, par exemple, d'une architecture souple autorisant l'intégration d'autres travaux dans ce domaine. Le problème de la gestion d'un comportement crédible pour le personnage virtuel incarnant le guide nous semble au contraire inévitable, l'aspect dynamique de l'environnement imposant des contraintes fortes sur ce comportement (apparition ou disparition d'objets à tout instant dans l'environnement réel que nous cherchons à décrire). La notion d'interruption proposées dans certains travaux nous semble de ce point de vue extrêmement intéressante. Il paraît néanmoins indispensable de disposer d'un contrôle précis sur ce mécanisme d'interruption, les facteurs déclenchant étant nombreux et moins contrôlés dans le cadre d'un environnement dynamique que dans celui d'une application d'interaction avec un utilisateur.

Finalement, nous avons développé les différents aspects de l'adaptation au travers des notions de personnalisation (adaptation *a priori*) et d'adaptivité (adaptation continue). Nous

avons décrit les différents objets et sources d'adaptation existants dans le cadre des guides virtuels. Dans le cadre de notre étude, nous nous concentrons sur l'adaptation par rapport à l'environnement. Celui-ci étant dynamique, il est nécessaire de prendre en compte son évolution lors de la sélection et la présentation d'informations aux visiteurs. Il est important de noter que cette adaptation par rapport à l'évolution de l'environnement ne peut se faire qu'à l'exécution (adaptivité). Il serait néanmoins intéressant de proposer une solution pouvant intégrer d'autres types et facteurs d'adaptation tels qu'une forme de personnalisation par rapport au visiteur. Cette présentation nous a également permis de mettre en avant la difficulté de la perception des différentes sources d'adaptation possibles et notamment l'environnement et l'utilisateur. Cet aspect est bien évidemment d'autant plus prégnant lorsque l'on s'intéresse à un environnement non contrôlé tel que celui considéré dans nos travaux.

Chapitre 3

Un guide virtuel autonome pour un environnement dynamique

3.1 Introduction

Dans les travaux que nous avons présentés on trouve la notion de visite dynamique (cf. 2.4), c'est à dire une visite dans laquelle le guide s'adapte aux actions de l'utilisateur ou à l'état de l'environnement. Cette notion de visite dynamique est différente de celle d'une visite au sein d'un environnement dynamique. Bien que la visite dynamique soit adaptée à un ensemble de facteurs (cf. 2.6), les objets sur lesquels la visite s'effectue restent généralement statiques. Le problème qui nous intéresse ici est sensiblement différent. Il concerne une visite se déroulant dans un environnement dynamique au sens où les informations fournies par le guide virtuel concernent des entités autonomes au sein de l'environnement. Ces entités possèdent un comportement propre qui n'est que peu ou pas prévisible.

Le guide virtuel et plus particulièrement les explications qu'il fournit (son discours) sont affectés par cette spécificité. Lorsque le guide explique au visiteur une entité particulière celle-ci peut à tout moment disparaître. De même à chaque instant, une nouvelle entité plus intéressante du point de vue d'un objectif pédagogique particulier peut faire son apparition. Dans un cas comme dans l'autre le guide virtuel doit être capable d'adapter son comportement. Il doit pouvoir interrompre une explication sur un sujet déterminé pour en déclencher une autre avant d'éventuellement revenir à l'explication interrompue. Ces changements doivent pouvoir être effectués de façon transparente pour le visiteur.

Dans ce chapitre nous détaillons notre proposition quand au processus de prise de décision d'un guide virtuel autonome capable de décrire un environnement dynamique (figure 3.1). Celle-ci repose sur deux points :

- ▷ le contenu des explications fournies au visiteur,
- ▷ la gestion du comportement du guide virtuel.

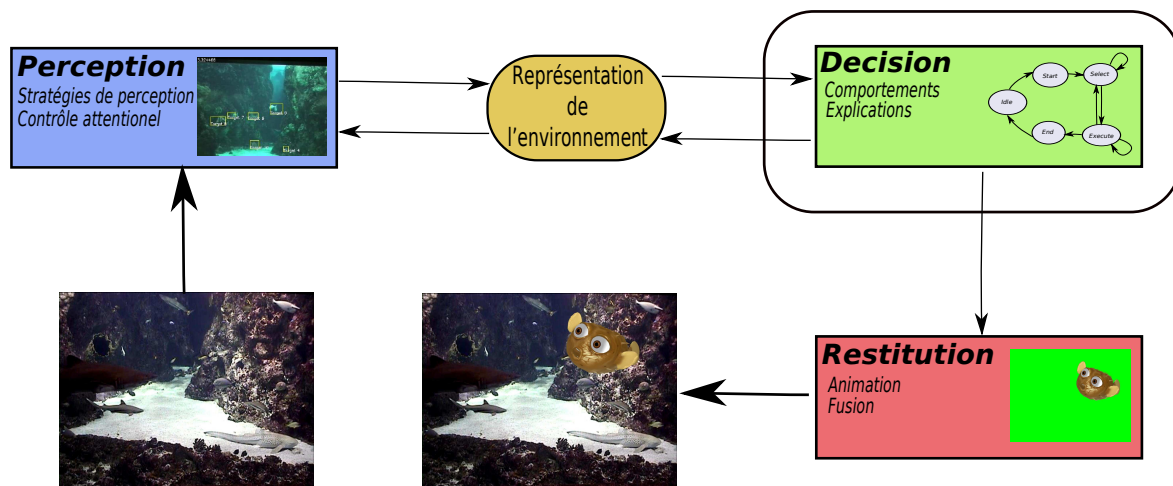


FIG. 3.1 – Positionnement du chapitre dans l'architecture générale de notre proposition.

Le premier point concerne la définition et la structuration des informations pouvant être présentées. Le second concerne l'intégration crédible du guide virtuel dans l'environnement réel et la présentation effective des informations.

3.2 Les explications

3.2.1 Introduction

Face aux contraintes imposées par l'environnement dynamique et les entités qui le peuplent, le guide virtuel doit pouvoir adapter son discours, les informations qu'il présente au visiteur, de manière appropriée. Il doit par exemple pouvoir s'arrêter pendant une explication parce qu'un événement important s'est produit dans l'environnement. Il doit aussi être en mesure de reprendre une explication là où elle a été interrompue. Bien entendu tous ces « sauts » dans l'ensemble du contenu existant ne doivent pas altérer la structure et la cohérence du discours effectué par le guide virtuel. Interrompre une explication particulière ou en reprendre une autre doit être fait de façon adaptée en évitant de s'arrêter au milieu d'une phrase par exemple et en faisant appel à des transitions appropriées.

Nous proposons de définir le contenu du discours sous la forme d'éléments simples organisés au sein d'*explications* pouvant être interrompues. Les interruptions donnent lieu à la coexistence de plusieurs processus d'explications simultanément lors de la visite, chacun pouvant être repris à l'endroit où il s'est arrêté. Un mécanisme de transitions permet d'assurer la cohérence lors du passage d'un processus d'explication à un autre.

Nous présentons ici les éléments permettant de définir du contenu pour le discours du guide virtuel. Nous nous intéressons au rôle de chacun de ces éléments et à la façon dont ils

sont utilisés pour définir l'espace des explications pouvant être fournies par le guide. Nous détaillons ensuite comment ces explications peuvent être exécutées et interrompues sous la forme de processus d'explication. Finalement, nous nous attardons sur les transitions et leur utilisation pour permettre le passage d'un processus d'explication à un autre.

3.2.2 La notion d'explication

La construction d'une explication s'effectue à partir de briques de base que nous nommons « éléments de discours » ou « éléments d'explication ». Un tel élément est une unité atomique d'explication. C'est le niveau de granularité le plus fin existant au sein du moteur de discours. L'interruption d'une explication n'est possible qu'avant ou après l'exécution d'un tel élément.

Une unité atomique de discours permet de définir la façon dont une information particulière doit être présentée au visiteur. Pour cela un élément de discours est constitué des propriétés suivantes :

- ▷ Un identifiant unique permettant de le référencer. C'est cet identifiant qui est utilisé au sein des explications et permet, par exemple, d'utiliser une même élément au sein de plusieurs explications.
- ▷ Les actions à effectuer par le guide lors de l'exécution de l'élément de discours.

Les actions que le guide virtuel peut effectuer dans l'environnement réel sont potentiellement illimitées. Nous avons décidé de nous concentrer sur la présentation d'informations au visiteur en mettant de côté les problèmes d'animation "bas-niveau" tels que le calcul de trajectoire ou l'animation du personnage virtuel. C'est un module de restitution spécialisé qui se charge de ces problèmes en présentant au processus de prise de décision une interface orientée sur la présentation de documents multimédias.

La gestion de la navigation y est effectuée de façon automatique à l'aide de techniques de *steering*[Reynolds 87]. Pour cela, le module d'animation se base sur les objets contenus dans la représentation de l'environnement. Il est possible de contrôler la navigation en affectant un mode de fonctionnement particulier. Selon le mode de navigation choisi, les mouvements du guide seront différents. Les déplacements ainsi que l'apparence générale du guide sont également affecté par un autre paramètre : l'état du guide virtuel. Bien que nous ne prenions pas en compte ici la gestion des émotions, cette information permet une plus grande variété dans l'animation du guide offrant ainsi plus de possibilités pour la présentation d'informations. Suivant l'état sélectionné, le guide va se déplacer plus ou moins rapidement et ses attitudes (animations du personnage qui sont également gérées automatiquement) seront différentes.

Les autres fonctionnalités du module d'animation sont focalisées sur la présentation d'information *via* la possibilité de lire différents types de documents et le déclenchement d'animations ponctuelles du personnage virtuel. La lecture de documents multimédias permet au guide de jouer des fichiers son, vidéo ou encore d'afficher des images qui seront intégrés dans l'environnement virtuel. Le déclenchement d'animations que nous qualifions ici de « ponctuelles » permet d'ajouter à la crédibilité du guide en lui donnant des attitudes/gestes

spécifiques à des moments déterminés. Ces différentes actions peuvent s'effectuer en parallèle, permettant par exemple le déclenchement d'un signe de salut à un moment précis de la lecture d'un fichier son, les différentes animations étant mixées ensemble au besoin. L'annexe B présente les détails des possibilités du module de restitution utilisé dans notre application.

Notre problématique n'étant pas axée sur les aspects animation et gestion des différentes modalités du communication, les actions à effectuer par le guide lors de l'exécution d'un élément de discours sont définies sous la forme d'un script. Ce script décrit un ensemble de contraintes associées à des commandes du module de restitution. Chaque contrainte consiste en un prédicat devant être vérifié pour que l'action associée soit exécutée. Ces contraintes peuvent être d'ordre temporel (date à laquelle l'action doit être déclenchée par rapport à la date de début d'exécution du script), des tests sur l'état interne du guide virtuel ou le contexte de la visite guidée (cf. 3.3.3). L'exécution du script ne se termine que lorsque toutes les actions qui le constitue ont été exécutées. La fin de l'exécution d'un script est notifiée au reste du système par le déclenchement d'un événement.

Un élément de discours correspond généralement à une ou plusieurs phrases permettant de transmettre une information spécifique de façon cohérente. Ces phrases sont pré-enregistrées et sont jouées comme n'importe quel autre type de document multimédia (images, vidéos). Des animations ponctuelles peuvent être ajoutées afin d'améliorer la crédibilité de la présentation des informations, par exemple en faisant effectuer au guide virtuel un clin d'oeil à un moment précis pendant la lecture d'un son.

La limitation principale des scripts d'animation est bien évidemment leur nature statique, les scripts sont définis une fois pour toute et ne peuvent en l'état être adaptés durant le déroulement de la visite. Pour palier à cette limitation il est possible de définir plusieurs scripts d'animation pour un même élément de discours. Le script devant être joué est sélectionné lors de l'exécution de l'élément de discours en fonction de conditions définies par les auteurs.

3.2.3 Graphe d'explication

Les éléments d'explication permettent de définir le contenu de base du discours de notre guide virtuel. Pour être capable d'effectuer des explications complètes et mettre en place une visite guidée ces éléments sont structurés les uns par rapport aux autres pour former une explication complète sous la forme d'un graphe orienté acyclique $G = (E, P)$ avec :

- ▷ E l'ensemble des nœuds de l'explication. Chaque nœud référence un élément de discours à l'aide de son identifiant unique ;
- ▷ P une relation de précédence entre deux éléments de discours spécifiques. Un couple (e_1, e_2) est un arc du graphe G si l'élément d'explication e_2 ne peut être exécuté qu'après l'exécution de l'élément d'explication e_1 .

Un graphe d'explication G est finalement décrit par :

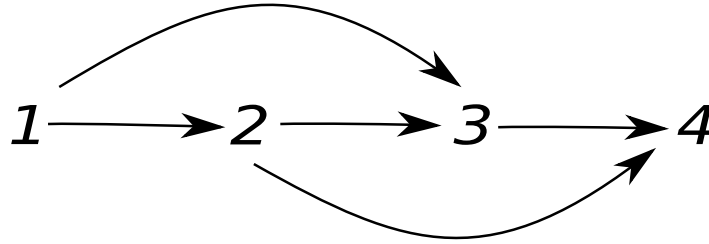
$$G = (E, P), E = \{e_1, e_2, \dots, e_{n-1}, e_n\}, P = \{p_1, p_2, \dots, p_n\}$$

où :

e_i est un élément d'explication atomique

p_i est une relation de précédence entre deux éléments d'explication e_i et e_j

$\forall p = (e_i, e_j) \in P, rang(e_i) < rang(e_j)$



1. Salut ! C'est moi Didon, le diodon. Je suis un poisson porc-épic et j'habite ce lagon de l'océan Pacifique.
2. Ici c'est mon domaine et je connais tout les poissons.
3. Tu vois, je suis entouré de requins, des petits et des gros avec de longues dents pointues...mais, qu'ils se méfient, moi aussi j'ai des piquants pointus.
4. Si tu veux, je t'emmène à la découverte de tout ce petit monde.

FIG. 3.2 – Exemple de graphe d'explication.

Un nœud e possède un ensemble de successeurs $S(e) = \{e_i, i \in \mathbb{N}\}$, plusieurs nœuds pouvant partager un même successeur. Les points de départ d'une explication sont définis par l'ensemble $D(G) = \{\{e_i\}, \neg \exists prec(e_i)\}$. Les nœuds terminaux à l'inverse sont définis par l'ensemble $F(G) = \{\{e_i\}, \neg \exists suc(e_i)\} \Leftrightarrow \{\{e_i\}, S(e_i) = \emptyset\}$.

La figure 3.2 présente un exemple d'un tel graphe d'explication. Celui-ci est constitué de quatre éléments de discours. Chacun correspond à une ou deux phrases pouvant s'enchaîner de différentes façons. Le choix d'un chemin particulier au sein de cette explication se fait lors de son exécution au sein d'un processus d'explication.

3.2.4 Exécution : processus d'explication

L'exécution d'une explication se fait par exécution successive d'éléments de discours constituant un chemin au sein du graphe de l'explication en question. Cette exécution que nous désignons sous le terme de « processus d'explication » est caractérisée par :

- ▷ un graphe d'explication G ;

- ▷ un éventuel support réel (entité visuelle présente dans la représentation de l'environnement) ou virtuel (image ou vidéo) ;
- ▷ le nœud courant du graphe d'explication : l'élément de discours en cours d'exécution ;
- ▷ l'ensemble des nœuds d'ors et déjà parcourus : le chemin $\{e_1, e_2, \dots, e_n\}$ où $e_1 \in D(G)$.

Un processus d'explication est considéré comme terminé lorsque le nœud e_m qui vient d'être exécuté est terminal, c.a.d. $e_m \in F(G)$. L'élaboration d'une explication consiste en l'exécution du script d'animation associé au nœud e_j lors du passage de la transition $e_i \rightarrow e_j$. Le choix du successeur e_j de e_i se fait selon une fonction $f : succ(e_i) \rightarrow e_j$ pouvant être spécifiée pour chacun des nœuds du graphe d'explication. Cette fonction se base sur le contexte de la visite (cf. 3.3.3) et associe à l'ensemble des successeurs de e_i un élément de discours spécifique.

Dans l'exemple de la figure 3.2 trois alternatives sont possibles lors de l'exécution du processus associé au graphe d'explication présenté : $1 \rightarrow 2 \rightarrow 3 \rightarrow 4$; $1 \rightarrow 3 \rightarrow 4$ ou $1 \rightarrow 2 \rightarrow 4$. Le choix de l'une ou l'autre de ces alternatives peut se faire de façon aléatoire afin d'apporter de la variété dans le discours du guide virtuel. Pour d'autres explications, ce choix pourra se faire en fonction du temps restant avant la fin de la visite ou encore d'un profil sélectionné par le visiteur.

Un processus d'explication peut être interrompu lors du passage d'un élément de discours à un autre, auquel cas il est suspendu est son état courant est sauvegardé. Cet état comprend :

- ▷ le nœud courant e_c du graphe d'explication G ;
- ▷ le chemin $C(G, E)$ caractérisé par la suite des éléments de discours exécutés jusqu'à présent pour le graphe d'explication G .

Ainsi, un processus d'explication Pr est défini comme suit :

$Pr = (G, S, e_c, C), G = (E, P), C = \{e_i, ; e_j\}$
<p>où :</p> <p>S est un éventuel support d'explication</p> <p>e_c un nœud du graphe G désigné sous le terme de nœud courant</p>

Du fait des interruptions, les explications sont structurées suivants trois ensembles :

- ▷ l'ensemble des explications terminées, c'est à dire l'ensemble des processus d'explication complètement traités pour lesquels il existe un chemin complet $C(G, E) = e_1, \dots, e_n; e_1 \in D(G), e_n \in F(G)$;
- ▷ l'ensemble des explication non encore abordées, c'est à dire l'ensemble des processus pour lesquels aucun chemin (même partiel) n'existe ($C(G, E) = \emptyset$) ;

- ▷ l'ensemble des explications suspendues, c'est à dire l'ensemble des processus pour lesquels il existe un chemin partiel $C(G, E) = e_1, \dots; e_n, e_1 \in D(G), e_n \neg \in F(G)$.

Ces trois ensembles auxquels s'ajoutent l'explication courante constituent l'état du discours du guide virtuel à un instant donné.

3.2.5 Transitions entre explications

L'interruption ou la reprise d'une explication de façon cohérente sont rendus possible par l'utilisation de transitions. Une transition est un ensemble d'éléments de discours équivalents, des alternatives, spécialisés en fonction d'une situation particulière (la passage d'un thème à un autre, la reprise d'une explication interrompue, *etc*).

La notion de transition est en fait légèrement plus générale que le terme ne le laisse entendre. Une transition, en plus de son utilisation en cas d'interruption ou de reprise, peut également servir lorsque l'on commence ou termine une explication, occupant ainsi le rôle d'introduction ou de conclusion.

Une transition est caractérisée par un prédicat défini par rapport au contexte de la visite permettant de spécifier ses conditions d'utilisation. La vérification de ce prédicat détermine la validité d'une transition particulière à un instant donné. Suivant le contexte de la visite (cf. section 3.3.3), plusieurs prédicats peuvent être vérifiés et de ce fait plusieurs transitions considérées comme valides. Lorsqu'on revient sur un scénario interrompu par exemple, il est tout à fait envisageable d'utiliser une transition de reprise. Si le scénario qui précède la reprise est arrivé à son terme, une transition de type conclusion peut être utilisée. Certaines transitions peuvent également être définies de façon plus ou moins spécifique pour une situation donnée (disparition d'une entité visuelle par exemple).

Situation	Transition(s)
Disparition de l'entité décrite	<ul style="list-style-type: none"> – Ho ! Il s'en va, laissons le, on le reverra plus tard. – Mais ? Ou est-il passé ? On le retrouvera sûrement plus tard.
Changement d'explication	<ul style="list-style-type: none"> – Bon...parlons d'autre chose. – Ah ! Voilà d'autres poissons que je connais bien. – Je cherche d'autre poissons. Ah voilà... – Voyons, voyons. Ah ! Voilà un autre poisson. – Continuons...
Début d'explication sur les platax	<ul style="list-style-type: none"> – Parlons un peu des platax. – Ces poissons marrons en forme de disque rayé sont des platax.
Retour sur une explication concernant les requins	<ul style="list-style-type: none"> – Ah ! Revoilà un requin. Ou en étais-je ?

TAB. 3.1 – Exemples de transitions et situations dans lesquelles elle doivent être utilisées.

Le tableau 3.1 présente différents exemples de transitions ainsi que les situations auxquelles elles sont destinées. Le choix de la (les) transition(s) à utiliser lors d'un changement ou du commencement d'une explication particulière se fait parmi l'ensemble des transitions valides à l'instant du changement à l'aide d'une fonction de sélection. L'exécution de la transition est à la charge du comportement courant du guide virtuel (cf. 3.4).

3.2.6 Bilan

La définition du discours du guide virtuel de façon modulaire au travers de l'ensemble des éléments de discours et de leur structuration au sein d'un graphe d'explication définissent un modèle d'exécution capable de gérer les interruptions pouvant intervenir dans le discours du guide virtuel. Ces interruptions sont inévitables du fait de l'aspect dynamique de l'environnement que le guide virtuel cherche à expliquer au visiteur. Elles donnent lieu à la coexistence de plusieurs processus d'explication en parallèle, un seul pouvant être exécuté à un instant donné : l'explication courante.

Le passage d'une explication à une autre est rendu possible par l'utilisation de transitions définies par rapport à une situation spécifique (apparition d'une entité visuelle plus intéressante, disparition de l'entité en cours d'explication, *etc*). Une transition peut être utilisée pour passer d'une explication à une autre mais également d'une façon plus générale sous la forme d'une introduction ou d'une conclusion. Le problème du choix de la (les) transition(s) devant être exécutée parmi l'ensemble des transitions valides est laissé au comportement du guide virtuel. Ce dernier point n'a pas été plus étudié, nous disposons néanmoins de quelques pistes quand à des mécanismes de combinaison entre les différentes transitions valides afin de construire une transition finale pouvant être utilisée. Ces mécanismes devront prendre en compte la spécificité des transitions par rapport à la situation ainsi que le « type » de situations pour lesquelles les transitions ont été définies (introduction, conclusion, changement d'explication, retour sur une explication interrompue, *etc*). Pour cela il est sans doute nécessaire de doter les transitions d'informations supplémentaires par rapport à ce que nous avons décrit dans la section 3.2.5.

L'exécution des explications entraîne la présentation des informations contenues dans un élément de discours *via* le déclenchement d'un script d'animation. Bien que l'utilisation de tels scripts peut sembler être une limitation, elle nous permet ici de contrôler précisément la façon dont une informations doit être présentée au visiteur. De plus la façon dont sont définis les éléments de discours offre la possibilité de remplacer de tels scripts par un moteur permettant de générer des commandes d'animation lors de l'exécution, autorisant ainsi la prise en compte, par exemple, d'émotions et la modification de la présentation d'informations en fonctions de celles-ci.

La structure du discours, le contenu des explications ainsi que les différentes transitions et situations sont définis par les auteurs. L'approche que nous proposons leur offre un contrôle précis sur la présentation d'informations et la structuration des différents éléments du discours (cf. annexe A). Le moteur que nous avons détaillé permet par exemple de définir différentes façons de présenter une même information, la forme appropriée étant sélectionnée en fonction du contexte de la visite lors de l'exécution.

Pour que la cohérence du discours soit réellement maintenue malgré les interruptions, les auteurs doivent structurer les explications de façon appropriée. Pour cela, il est important d'observer quelques « règles » :

- ▷ les éléments de discours doivent être courts (une ou deux phrases) afin d'autoriser les interruptions le plus souvent possible. Des exemples d'éléments de discours peuvent être trouvés dans la section 6.3.2.
- ▷ les éléments d'introduction et de conclusion d'une explication ne doivent pas y être intégré, ce type d'élément doivent être définis en tant que transition (cf. 3.2.5)
- ▷ une information précise pour un élément de discours unique permet de faciliter la gestion des interruptions et d'assurer une meilleure cohérence lors de l'exécution (retour sur une explication par exemple). Les phrases suivantes par exemple devront plutôt être placées dans un même élément de discours :
« Je trouve que le jeune a un peu l'apparence d'une chauve-souris. D'ailleurs en anglais, ces poissons sont appelé *batfish* : *bat* comme *Batman* et *fish* comme poisson ! »

3.3 La visite guidée : une succession d'explications

3.3.1 Introduction

La visite guidée est une succession d'explications permettant de présenter des informations spécifiques au visiteur au travers du contrôle des actions du guide virtuel.

Compte tenu des contraintes de l'environnement dynamique, le chemin au sein des données de discours qui définit la visite guidée ne peut être prévu à l'avance. Il est nécessaire de construire la visite au fur et à mesure de son déroulement. La visite guidée doit malgré tout être structurée et nous devons permettre aux auteurs d'avoir une influence sur le processus de construction. Ce sont eux qui doivent fournir au guide les connaissances nécessaires sur la visite. Ceci est rendu possible grâce à l'ajout d'informations sémantiques aux explications au travers de la notion de sujet et par la définition d'un ensemble d'experts permettant d'exprimer différents points de vue sur le choix des explications à présenter aux visiteurs.

3.3.2 Notion de sujet

Un sujet caractérise une explication au travers des propriétés suivantes :

- ▷ un ensemble de thèmes (mots-clés) qui décrivent le contenu du sujet ;
- ▷ un score initialement nul qui détermine la pertinence de l'explication par rapport au contexte courant de la visite ;
- ▷ une référence sur l'explication caractérisée par le sujet.

Ainsi les sujets permettent de caractériser les explications de façon transversale à travers une liste de thèmes. Dans le cadre de notre application nous avons définis notamment les sujets suivants (les sujet sont présentés en ligne et les différents thèmes en colonnes) :

Sujet \ Thème	Requin	Platax	Naso	Reproduction	Habitat
Platax-1	-	X	-	-	X
Naso-1	-	-	X	X	-
Naso-1	-	-	X	-	X
Requin-1	X	-	-	-	X

Une liste exhaustive des sujet et thèmes associés est présentées dans la section 6.2.3.

Plusieurs explications peuvent éventuellement être associées à un même sujet. Bien que cette possibilité existe, elle n'est pas exploitée et actuellement il n'existe qu'une relation directe entre un sujet et une explication spécifique (*i.e.* il y a autant de sujets que d'explications).

Alors que les éléments de discours permettent de décrire la présentation des informations et que les explications et transitions permettent de structurer ces différents éléments, les sujets sont utilisés par le processus de prise de décision de notre agent autonome afin de sélectionner les explications qui seront présentées au visiteur. Dans la suite de ce chapitre nous emploierons le terme sujet pour parler d'un processus d'explication et des informations qui lui sont associées.

3.3.3 Contexte de la visite

Le choix du sujet courant de la visite se fait par rapport à un contexte que nous allons maintenant expliciter. Le contexte de la visite contient toutes les informations permettant au processus de prise de décision de choisir les explications à fournir au visiteur. Ce contexte est constitué des éléments suivants :

- ▷ le temps écoulé depuis le début de la visite ;
- ▷ un historique de la visite qui contient l'ensemble des sujets abordés ainsi que des informations sur leur éventuelle interruption au cours du temps ;
- ▷ l'ensemble des sujet disponibles, c'est à dire les sujets existant n'ayant pas encore été totalement présentés au visiteur ainsi que les sujets dont le processus d'explication est suspendu.
- ▷ l'ensemble des transitions valides à un instant donné. La validité d'une transition particulière est déterminé par rapport au reste du contexte de la visite après le choix du sujet à présenter au visiteur.
- ▷ un sujet courant (processus d'explication et les informations qui lui sont associées) ;
- ▷ une liste d'événements ;
- ▷ une représentation de l'environnement provenant du module de perception.

Historique

L'historique de la visite stocke des informations sur l'évolution de la visite guidée sous la forme d'une série d'événements concernant les explications. Nous considérons quatre types d'événements :

- ▷ le début d'une explication (`:begin`) ;
- l'interruption d'une explication (`:interrupted`) ;
- la reprise d'une explication (`:resume`) ;
- la fin d'une explication (`:end`) .

Les événements associés à chaque sujet (processus d'explication) sont stockés en même temps que l'identifiant du sujet et une date par rapport au début de la visite guidée. Cet historique est mis à jour chaque fois que le sujet courant de la visite est réévalué, ce qui inclut le cas où un sujet interrompu est sélectionné à nouveau immédiatement après son interruption.

Ces informations nous permettent à tout instant de connaître :

- ▷ l'historique d'un sujet/thème spécifique ;
- ▷ l'état courant d'un sujet (non abordé, commencé, terminé) ;
- ▷ si un sujet/thème spécifique a été interrompu ;
- ▷ depuis quand un sujet/thème est traité ;
- ▷ le temps total passé sur un sujet/thème depuis le début de la visite.

De telles informations sont utilisées non seulement pour le déroulement de la visite guidée (sélection de l'explication courante) mais également pour tester la validité des transitions à un instant donné. La validité de chacune des transitions disponibles est vérifiée chaque fois qu'un changement de sujet (processus d'explication) intervient.

Évènements

Les événements décrivent une modification importante de l'environnement. Nous distinguons plusieurs types d'événements :

- ▷ les événements internes. Cela concerne l'écoulement du temps imparti pour la visite, la notification de la fin d'un script d'animation ou encore l'interruption d'une explication.
- ▷ les événements externes. Cela concerne principalement les événements se produisant au sein de l'environnement à expliquer comme par exemple l'apparition ou la disparition d'une entité visuelle.

Le processus de prise de décision du guide virtuel consiste principalement en la sélection de l'explication courante et en la réévaluation de cette explication lorsqu'un événement particulier intervient. Nous considérons notamment les événements suivants comme pouvant entraîner une réévaluation du sujet courant :

- ▷ la fin de l'explication courante ;
- ▷ l'apparition d'une nouvelle entité visuelle dans la représentation de l'environnement ;
- ▷ la disparition d'une entité support de l'explication courante.

Environnement

La représentation de l'environnement quant à elle comporte un certain nombre d'entités visuelles auxquelles sont associées des propriétés parmi lesquelles :

- ▷ une position dans l'espace ;
- ▷ une taille (volume englobant) ;
- ▷ un type.

D'autres propriétés existent et le détail de la représentation de l'environnement est présenté dans la section 5.3.2.

Cette représentation peut être interrogée à tout moment par le guide virtuel à l'aide de requêtes à la manière d'une base de données. Ces requêtes permettent de spécifier des contraintes sur les différentes propriétés des entités visuelles qui sont intéressantes dans la représentation. L'exécution de chaque requête permet de récupérer une liste d'entités visuelles qui satisfont aux contraintes spécifiées.

```
(get-visual-entities (or (and (type 'Shark')
                              (distance < 100))
                        (approaching)))
```

LISTING 3.1 – Exemple d'interrogation de la représentation de l'environnement

Le listing 3.1 montre un exemple d'une telle requête. Les contraintes utilisées permettent de récupérer les objets qui approchent de la position courante du guide virtuel (au moment de la requête) ainsi que les objets de type "Shark" se trouvant à une distance inférieure à 100 unités de cette même position. Le fonctionnement de ces requêtes est détaillé dans la section 5.4.4.

Les mécanismes qui permettent la création et la mise à jour de la représentation ne sont pas décrits ici mais nous demandons au lecteur de considérer que ceux-ci assurent un contenu cohérent par rapport à l'état réel de l'environnement même si ce contenu peut s'avérer incomplet (certaines entités peuvent ne pas être identifiées ou détectées par exemple). Le détail des mécanismes de perception visuelles sont présentés dans le chapitre 5.

Algorithme 3.3.1: SELECTIONDESUJET(*allExperts*, *allSubjects*, *tourContext*)

```

– allExperts: ensemble des experts actifs
– allSubjects: ensemble des sujets disponibles
– tourContext: contexte de la visite (environnement, etc)
/* Chaque expert vote sur la liste de sujets renvoyant une liste de votes. */
for all expert ∈ allExperts
  do votes[expert] ← VOTE(expert, sujets, tourContext)
/* Les votes de chaque expert sont combinés dans une liste finale. */
/* Chaque vote peut prendre l'une des valeurs : 0, 1, VETO */
for each s ∈ allSubjects
  do {
    finalVotes[s] ← 0
    for each e ∈ allExperts
      do {
        if finalVotes[s] ≠ VETO
          then if votes[e][s] ≠ VETO
            then finalVotes[s] ← finalVotes[s] + votes[e][s]
          else finalVotes[s] ← VETO
      }
  }
/* On choisit de façon aléatoire un sujet parmi ceux possédant le plus */
/* grand nombre de votes. */
return (RANDOM(MAXIMUMSCORES(finalVotes)))

```

3.3.4 Sélection des explications

La sélection des explications est effectuée par un ensemble d'experts. Chaque expert affecte un score aux explications en fonction d'un point de vue. Chaque fois qu'il est nécessaire de choisir un sujet, l'ensemble des sujets suspendus ainsi que celui des sujets non encore commencés sont considérés par les différents experts lors d'une procédure de vote. A chacun de ces sujets est associé une score définissant sa pertinence par rapport au contexte de la visite en tant que sujet courant. Ce score est nul au début de la procédure de sélection puis est modifié au fur et à mesure de cette procédure par les différents experts actifs. Nous allons détailler plus loin ce qu'est un tel expert et son fonctionnement. Un fois l'ensemble des experts exécutés, nous disposons d'une liste contenant les sujets disponibles ordonnés suivant l'influence des différents experts impliqué dans la procédure de sélection. A partir de cette liste, le sujet possédant le score le plus élevé est sélectionné. Si plusieurs sujets possèdent un score identique, le sujet devant être développé est choisi de façon aléatoire parmi ces sujets¹. Ce processus de sélection est résumé dans l'algorithme 3.3.1.

Le rôle d'un expert est d'apporter un « point de vue » sur la pertinence par rapport

¹ Ce critère de sélection peut être redéfini en fonction de l'application envisagée.

au contexte de la visite de tel ou tel sujet parmi la liste des sujets disponibles en tant que nouveau sujet courant. Pour cela un expert se voit présenté la liste de tous les sujets disponibles, chacun possédant alors un score nul. L'expert passe alors en revue l'ensemble de ces sujets afin de modifier ce score suivant ce pourquoi il a été défini. On peut distinguer différents types d'experts parmi lesquels :

- ▷ des experts se basant sur des connaissances *a priori* comme un ordre de préférence défini par les auteurs par exemple ;
- ▷ des experts se basant sur le contenu de l'environnement, préférant par exemple les sujets en rapport avec les entités visuelles présentes dans la représentation, ou encore des experts tenant compte de la distance des entités visuelles par rapport à la position courante du guide virtuel ;
- ▷ des experts se basant sur l'historique de la visite, favorisant par exemple les thèmes non encore abordés, les sujets suspendus ou encore les sujets sur lesquels le guide n'a pas passé suffisamment de temps jusqu'à présent.

C'est au travers de cette notion d'expert que le cours de la visite guidée est influencé et le sujet courant de celle-ci sélectionné. Elle permet aux auteurs de définir des tendances quand au déroulement de la visite guidée (du point de vue des informations à transmettre) sans pour autant définir un plan fixe pour celle-ci².

Pour chaque sujet, un expert peut effectuer l'une des actions suivantes :

- ▷ affecter une voix en faveur du sujet en question ;
- ▷ laisser inchangé le nombre de voix (*i.e.* ne rien faire) ;
- ▷ utiliser un droit de veto.

Le score associé à un sujet donné consiste en la somme des voix qui lui ont été attribuées par les différents experts. Le droit de veto consiste en la possibilité pour un expert d'empêcher la sélection d'un sujet donné, une telle action ayant pour effet de retirer le sujet en question de la liste des sujets disponibles lors du choix final effectué après la phase de vote. Un tel mécanisme permet, en plus du contrôle souple offert par le système de score, de donner un contrôle fort aux auteurs qui définissent les experts dans certaines situations. Cela peut être le cas par exemple d'un expert voulant s'assurer que l'on ne parle que des sujets pour lesquels il existe une entité visuelle appropriée dans la représentation de l'environnement. Les experts sont exécutés en parallèle et le score total n'est calculé qu'à la fin de cette exécution. Ainsi l'ordre n'influence pas le processus de sélection et chaque expert se voit présenter la totalité des sujets disponibles au départ du processus de sélection³.

Les experts peuvent être activés ou désactivés au cours de la visite. Ceci peut se faire au sein d'un comportement (cf. 3.4) ou directement au sein un élément de discours à l'aide des actions associées à chacun d'entre eux.

² Ce qui n'est pas possible étant donné l'aspect dynamique de l'environnement à expliquer.

³ Ceci a son importance, certains experts pouvant nécessiter une connaissance sur les sujets disponibles afin de faire leur choix.

3.3.5 La visite guidée

La sélection et la présentation d'une succession de sujets constitue finalement une visite guidée complète. Cette succession d'explications et de transitions, d'interruptions, de changements et de reprises décrit un chemin au sein de l'ensemble du discours défini par les auteurs (figure 3.3). L'exécution de ce chemin consiste en l'exécution successive de différents éléments de discours et de transitions permettant de lier les éléments de discours appartenant à des sujet différents.

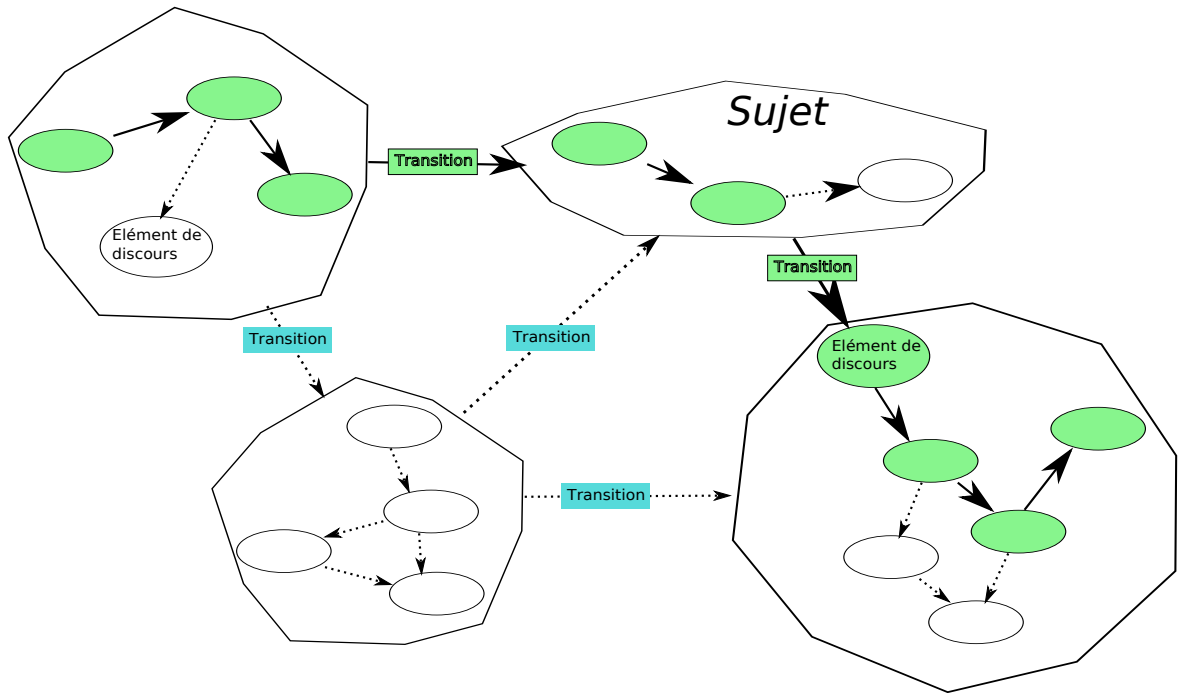


FIG. 3.3 – Vue d'ensemble des données de discours et d'une instance de visite guidée (chemin au sein de ces données).

Sur la figure 3.3 on peut voir un exemple simple d'une visite guidée. Ici quatre sujets existent, chacun étant associé à une explication composée de plusieurs éléments de discours. Cinq transitions permettent le passage de certains sujets à d'autres. Le déroulement de la visite guidée est représenté par les éléments (éléments de discours et transitions) colorés. Celle-ci implique ici trois des quatre sujets possibles, certains éléments de discours des explications associées à ces sujets n'étant pas exécutés. Un exemple d'interruption simple est présenté pour le sujet dont l'explication n'est pas exécutée jusqu'à un nœud terminal. Les interruptions plus complexes ne sont pas représentées sur ce schéma afin de le garder relativement simple, des exemples d'exécution plus complexes sont présentés dans la section 6.3.

3.3.6 Bilan

La notion de sujet que nous avons présentée permet d'ajouter des informations sémantiques aux explications (contenus) définis par les auteurs de la visite guidée. L'enchaînement de ces sujets (processus d'explications « augmentés » d'informations sémantiques) donne naissance à la visite guidée à proprement parler.

La sélection du sujet courant de la visite est effectuée par un ensemble d'experts spécialisés suivant différents points de vue. Ces experts affectent un score à chacun des sujets pouvant être candidat au rôle de sujet courant de la visite, c'est à dire l'ensemble des sujets suspendus et des sujets n'ayant pas encore été traités. Le sujet avec le score le plus élevé étant sélectionné pour être le sujet courant de la visite.

Ce processus de sélection s'effectue par rapport à un contexte qui définit l'ensemble des paramètres pouvant affecter le déroulement de la visite guidée. Ce contexte comprend notamment : l'ensemble des sujets disponibles, la représentation de l'environnement provenant du module de perception ainsi que l'ensemble des transitions valides à un instant donné.

L'utilisation de la notion d'expert pour déterminer le déroulement de la visite guidée permet la définition de « tendances » par les auteurs du discours du guide virtuel tout en laissant le choix final au guide virtuel au fur et à mesure de la visite guidée. Le contexte de la visite ne pouvant être prévu à l'avance toute planification *a priori* des sujets devant être traités est impossible.

Néanmoins ce mécanisme ne définit que la sélection des explications devant être présentées au visiteur. La mise en place de la visite guidée complète inclut la gestion du comportement du guide virtuel au sein duquel ce processus de sélection et l'exécution du processus d'explication sélectionné vont être mêlés.

3.4 Comportement du guide virtuel

3.4.1 Introduction

Les mécanismes et structures de données que nous avons présentés jusqu'à présent nous offrent la possibilité de définir le contenu du discours de notre guide virtuel ainsi que de sélectionner parmi ce contenu ce qui doit être présenté à un instant donné en fonction du contexte de la visite. Néanmoins, cela reste insuffisant pour permettre au guide virtuel d'effectuer réellement la visite guidée.

Une visite guidée complète implique le contrôle des actions du guide de façon à naviguer au sein de l'environnement à décrire, sélectionner une explication courante et présenter effectivement les éléments qui constituent cette explication. Ce sont ces différentes fonctions que nous regroupons sous le terme de comportement.

La particularité de l'environnement que notre guide cherche à expliquer à ici une influence importante sur cette notion de comportement. L'aspect dynamique et l'impossibilité de

prévoir les événements qui vont s'y produire nous oblige à mettre un place des mécanismes d'interruption permettant au guide virtuel de modifier son comportement lorsque cela s'avère nécessaire.

3.4.2 Notion de comportement

Le comportement du guide virtuel est régi par un automate hiérarchique séquentiel. La notion de parallélisme dans l'exécution du comportement n'est ici pas nécessaire du fait de la nature séquentielle des actions que le guide doit effectuer.

Cet automate hiérarchique est défini par :

- ▷ un ensemble d'états, ces états pouvant être de type élémentaire ou composite ;
- ▷ un ensemble de transitions, une transition étant un triplet (e_1, e_2, p) où e_1 et e_2 sont deux états et p est un prédicat portant sur le contexte de la visite.

Les états élémentaires décrivent des commandes pour le module de restitution et/ou la progression dans l'élaboration de l'explication courante. Les états composites sont constitués d'un ensemble de sous-automates pouvant être combinés de façon séquentielle ($Seq([a_1, \dots, a_n])$) ou concurrente ($Alt([a_1, \dots, a_n])$). Le premier type de combinaison conduit à l'exécution successive des automates a_1, a_2 jusque a_n . Le second type de combinaison conduit à la sélection d'un automate particulier dans l'ensemble $\{a_1, \dots, a_n\}$ au moyen d'une fonction $f : [a_1, \dots, a_n] \rightarrow a_i$. Cette fonction peut être un simple choix aléatoire ou un choix plus complexe se basant sur le contexte de la visite. L'exemple de la figure 3.4 présente un automate comportant à la fois des états composites (*Attente* et *Explication*) et élémentaires (*Sélection de sujet*) au premier niveau de la hiérarchie. Les sous-automates des deux états composites sont quant à eux constitués uniquement d'états élémentaires.

La transition (e_c, e_s, p) de l'état courant e_c du comportement vers un autre état e_s est possible quand :

- ▷ e_c est un état élémentaire et le prédicat p associé à la transition est vérifié ;
- ▷ e_c est un état composite séquentiel $Seq([a_1, \dots, a_n])$ dont le sous-automate a_n est dans son état final.
- ▷ e_c est un état composite concurrent $Alt([a_1, \dots, a_n])$ dont le sous automate sélectionné a_c est dans son état final.

La définition du comportement du guide virtuel, son « cycle de vie » passe par la définition d'un automate hiérarchique. La figure 3.4 présente un exemple d'un tel automate composé de trois états principaux :

- ▷ *Attente* – un état composite combinant ses sous-automates de façon séquentielle ;

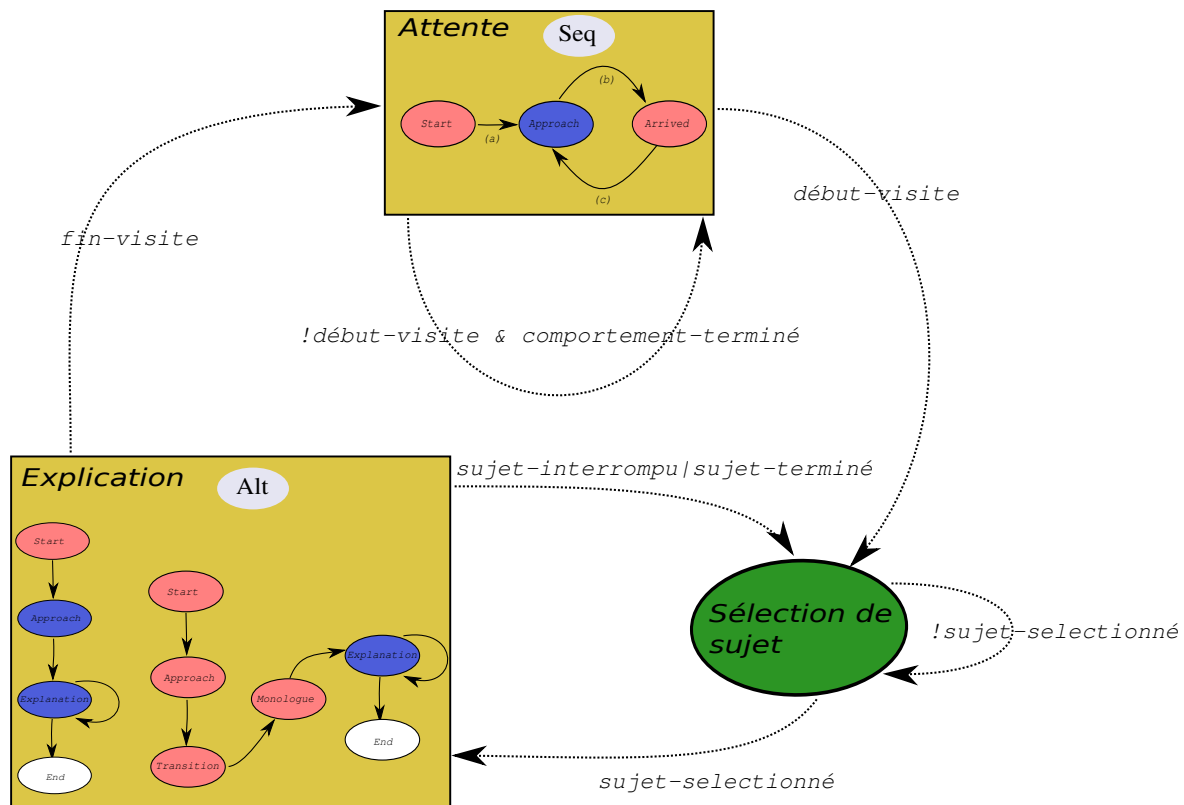


FIG. 3.4 – Exemple d'automate hiérarchique pouvant être utilisé dans le cadre du guide virtuel.

- ▷ *Sélection de sujet* – un état élémentaire ;
- ▷ *Explication* – un état composite combinant ses sous-automates de façon concurrente.

Cet automate permet la mise en place d'un comportement simple où le guide virtuel alterne la visite guidée avec des phases d'attente.

Dans l'état *Attente* un comportement est sélectionné par ceux disponibles puis exécuté. Cette exécution continue tant que le comportement n'est pas arrivé à son terme ou interrompu. La transition vers l'état suivant est possible lorsque l'événement *début-visite* est détecté.

L'automate passe alors dans l'état *Sélection de sujet* qui met en oeuvre le processus de sélection que nous avons décrit dans la section 3.3.4. Il reste dans cet état tant qu'aucun sujet n'est ressorti vainqueur du processus de sélection⁴.

Une fois un sujet sélectionné, l'automate passe dans l'état d'*Explication* qui permet l'exécution du sujet sélectionné. Ici un comportement va être sélectionné en fonction du contexte de la visite et du sujet qui vient d'être choisi. Ce comportement va alors être exécuté jusqu'à ce que le processus d'explication en cours soit terminé ou qu'un événement nécessitant une

⁴ Suivant la configuration d'expert mise en oeuvre il est possible qu'aucun sujet ne ressorte vainqueur du processus de sélection notamment à cause du droit de veto. Une telle situation n'est pas prise en compte dans notre exemple mais doit bien évidemment faire l'objet d'une attention toute particulière lors de la création de l'application souhaitée.

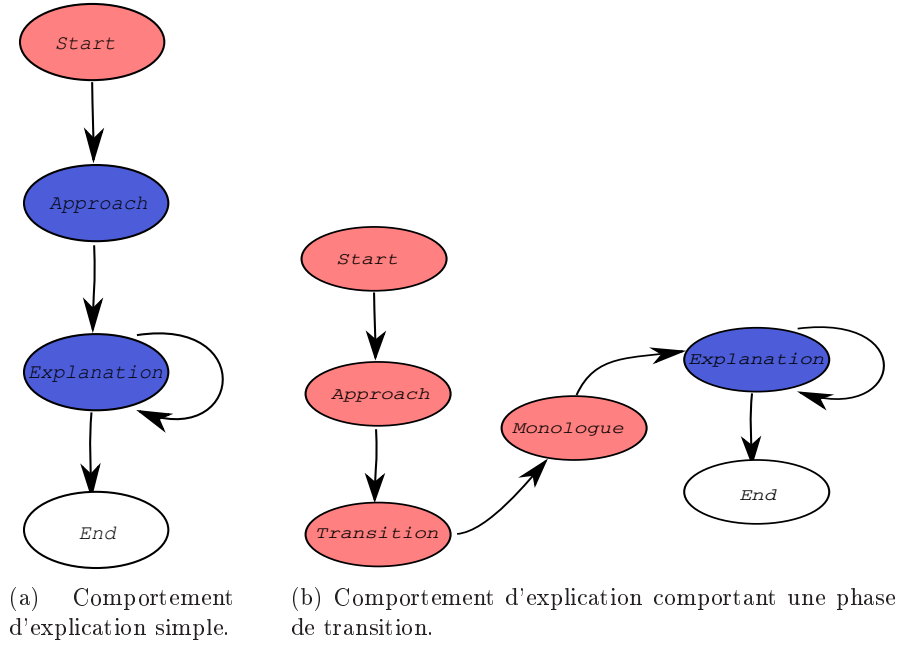


FIG. 3.5 – Exemples de sous-automates interruptibles

réévaluation du sujet courant intervienne. Dans les deux cas, l'automate va alors retourner dans l'état de *Sélection de sujet*. Si un événement de fin de visite intervient (temps imparti écoulé par exemple), l'automate va alors retourner dans son état initial d'*Attente*.

3.4.3 Exécution et gestion de l'interruptibilité

A tout instant l'état du guide virtuel est défini par l'état élémentaire e_c dans l'automate hiérarchique définissant son comportement. Cet état élémentaire courant peut appartenir à l'automate hiérarchique ou à l'un des sous-automates d'un état composite. Pour faire face à l'aspect dynamique de l'environnement dans lequel évolue notre guide virtuel, il est nécessaire de disposer d'un mécanisme lui permettant de modifier (interrompant éventuellement l'action en cours) cet état en fonction du contexte de la visite de façon appropriée, c'est à dire en maintenant à la fois sa crédibilité et la cohérence de la visite guidée qui est éventuellement en train d'être élaborée. Lorsque l'état courant e_c appartient à un état composite, la notion d'interruption permet de définir un "raccourci" qui, dans certaines conditions, va permettre d'interrompre prématurément le sous-automate en cours d'exécution. Une telle interruption du comportement peut être nécessaire lorsqu'une entité visuelle utilisée comme support d'une explication disparaît ou encore lorsqu'une nouvelle entité fait son apparition, le sujet courant de la visite nécessitant alors une réévaluation.

Pour que les contraintes de cohérence du discours et de crédibilité du comportement soient satisfaites, il est nécessaire que ces interruptions ne puissent intervenir qu'à des moments précis du déroulement de la visite déterminés lors de la définition des comportements. Il faut par exemple que le guide termine sa phrase avant de parler d'autre chose ou encore s'assurer qu'au

moins un élément de discours est exécuté après une transition afin d'assurer la cohérence de la visite guidée.

La gestion « fine » de telles situations d'interruption est rendue réalisable par la possibilité de marquer un état élémentaire du comportement comme étant interruptible ou non. Seuls ces états peuvent être marqués ainsi : ce sont les seuls états pouvant être considérés comme atomiques dans l'automate hiérarchique décrivant le comportement du guide virtuel. La gestion des interruptions n'est effective que pour les sous-automates du comportement, leur déclenchement entraînant alors le retour au niveau supérieur de l'automate hiérarchique.

La figure 3.5 présente des exemples de sous-automates faisant appel à des états interruptibles. Le premier exemple présente un comportement d'explication simple. Ici le guide s'approche de l'entité visuelle qu'il cherche à décrire. Un fois suffisamment proche, il commence son explication. Il peut être interrompu soit lorsqu'il est en cours d'approche (si une entité visuelle plus intéressante est repérée par exemple), soit lorsque le script d'animation de l'élément d'explication courant est terminé. Ceci se traduit dans la définition du sous-automate par le fait que les états élémentaires *Approach* et *Explanation* sont marqués comme étant interruptibles. Sur le second exemple, un comportement d'explication contenant l'exécution d'une transition est présenté. Utilisé lors du passage d'un sujet à un autre, ici le guide s'approche de l'entité visuelle à décrire puis, une fois arrivé à une distance acceptable, il commence sa transition avant de commencer son explication. On joue ici sur le contrôle de l'interruptibilité pour s'assurer que l'on effectue la transition ainsi qu'un minimum d'explication avant de pouvoir être interrompu dans le but de maintenir la cohérence du discours du guide virtuel. Ainsi, seul l'état *Explanation* est ici interruptible, obligeant l'état courant e_c à passer successivement par les états *Approach*, *Transition* et *Monologue* correspondants aux différentes étapes que nous avons décrites précédemment.

L'interruptibilité de l'état courant e_c est testée lors de chaque évaluation de cet état. L'interruption du sous-automate en cours d'exécution ne peut s'effectuer que dans des conditions précises associées à l'état composite chargé de l'exécution du sous-automate. Dans l'exemple de la figure 3.4, on trouve notamment les conditions suivantes :

- ▷ le début de la visite ;
- ▷ la fin de la visite ;
- ▷ la fin du processus d'explication courant ;
- ▷ l'interruption du processus d'explication courant pour l'une des raisons suivantes :
 - ◊ apparition d'une entité dans la représentation de l'environnement ;
 - ◊ disparition d'une entité dans la représentation de l'environnement.

Au cours de l'exécution, lorsque les conditions susceptibles de déclencher une interruption se produisent et que l'état courant est marqué comme étant interruptible, une demande d'interruption est déclenchée. Généralement les demandes d'interruption sont satisfaites immédiatement lorsque l'état est interruptible. Il est néanmoins possible de définir des conditions supplémentaires sur le contexte de la visite afin de spécialiser l'interruptibilité

d'un état (position du guide par exemple). Un tel mécanisme permet de limiter le nombre d'états lors de la définition du comportement⁵.

Si aucune demande d'interruption n'intervient, le sous-automate courant du guide progresse « normalement » de l'état élémentaire courant à l'un de ses états suivants. L'évolution du comportement du guide virtuel est définie par la succession des différents états de l'automate hiérarchique au cours de son exécution.

Un sous-automate se différencie de l'automate principal par le fait qu'il doit posséder obligatoirement les états suivants :

- ▷ *START* – état dans lequel le sous-automate démarre, en cas d'interruption l'état courant de l'automate est mis à jour pour correspondre à cet état de départ
- ▷ *END* – état final du sous-automate. Lorsqu'il arrive dans cet état, le sous-automate redonne le contrôle de l'exécution à l'automate de niveau supérieur
- ▷ *INTERRUPT* – état dans lequel le sous-automate se retrouve lorsqu'il reçoit une demande d'interruption de la part de l'automate de niveau supérieur, ici aussi le passage par cet état entraîne le passage du contrôle de l'exécution au niveau supérieur

Ces états obligatoires sont déclarés de façon automatique lors de la définition d'un sous-automate comme le présente l'annexe A.

3.4.4 Bilan

Nous avons proposé d'utiliser un automate hiérarchique séquentiel pour la gestion du comportement de notre guide virtuel. Cet automate permet de contrôler les actions de l'agent autonome dans son environnement lui permettant de s'y intégrer de façon crédible et de présenter des informations appropriées au visiteur. C'est lui qui se charge de coordonner le choix et l'exécution du processus d'explication courant en fonction du contexte de la visite guidée.

Un automate est composé d'états et de transitions, chaque état pouvant être élémentaire ou composé d'un ensemble de sous-automates. L'exécution du comportement est séquentielle, l'exécution d'un nœud composite pouvant s'effectuer de façon séquentielle ou concurrente.

La mise en place de la notion d'interruptibilité au niveau des états élémentaires de cet automate permet une gestion fine des changements pouvant intervenir du fait de l'environnement dynamique.

La définition du comportement précis du guide dépend de l'application envisagée, néanmoins une telle approche présente l'avantage d'offrir aux auteurs un contrôle fin sur les actions du guide et le déroulement de la visite au travers de la définition simple d'automates (cf. annexe A).

⁵ Ce mécanisme n'est pas nécessaire pour autant et ne constitue qu'une facilité à la disposition des auteurs lors de la définition du comportement du guide virtuel.

3.5 Déroulement d'une visite

Partant des multiples aspects que nous avons mis en place tout au long de ce chapitre, à savoir les processus d'explication et sujets ainsi que le comportement du guide virtuel, le déroulement complet d'une visite guidée est déterminé par l'évolution conjointe de l'état du guide virtuel (l'état au sein de son comportement) et du processus d'explication (sujet) courant du discours.

D'un point de vue général, la visite guidée consiste en le choix d'un sujet à exposer au visiteur et d'un sous-automate permettant d'exécuter ce sujet de façon appropriée. Cet automate est exécuté jusqu'à son terme ou jusqu'à ce qu'il soit interrompu (il doit pour cela autoriser les interruptions en au moins l'un des états élémentaires qui le constitue). Suivant le type d'interruption, le processus peut recommencer et un nouveau sujet être sélectionné ou la visite peut simplement se terminer.

Les types d'interruption et la marche à suivre en fonction de ce celles-ci dépendent de l'application et de la structure souhaitée pour la visite guidée. Dans notre application (cf. section 6.2), le cycle de vie du guide virtuel est structuré autour de la notion de sessions d'explication qui correspondent à une visite guidée d'une durée limitée définie par avance. Ces sessions alternent avec des phases d'attente au cours desquelles le guide évolue de façon crédible dans l'environnement sans délivrer d'informations au visiteur. Ces deux phases sont liées entre elles par des phases d'introduction et de conclusion permettant respectivement de le mettre à l'aise avec le système et de lui notifier la fin de la visite. Dans le cadre d'une autre application, nous pourrions imaginer que le guide effectue la visite par étapes, lorsque le visiteur se trouve face à quelque chose d'intéressant par exemple, intercalant des phases d'attentes au sein de la même visite guidée.

Les événements pouvant interrompre la visite ainsi que la marche à suivre lorsqu'une telle interruption a lieu peuvent bien évidemment varier sensiblement suivant le type d'application envisagée. Dans le premier cas, la visite sera interrompue lorsqu'un événement intéressant se produit dans l'environnement (apparition d'une entité visuelle ou disparition de l'entité visuelle en cours de description par exemple) ou lorsque la session d'explication se termine (le temps alloué est écoulé). L'occurrence d'un événement intéressant dans l'environnement déclenchera une réévaluation du sujet courant pour déterminer la suite de la visite alors que la fin de la session d'explication entraînera le déclenchement d'une phase de conclusion et le retour dans une phase d'attente du guide virtuel. Dans le second cas, la visite sera interrompue par le déplacement du visiteur lorsqu'il quitte le point de visite courant. Dans ce cas, le guide entrera dans une phase d'attente jusqu'à ce que le visiteur se trouve à un nouveau point de visite intéressant.

Malgré ces différences en fonction de l'application et de la structure que l'on souhaite donner à la visite guidée, il est possible d'identifier plusieurs types distincts de comportements et de phases dans ces comportements. On peut tout d'abord distinguer les comportement d'attente, au cours desquels le guide virtuel continue d'être intégré dans l'environnement réel de façon crédible mais ne délivre aucune explication au visiteur, et les comportement d'explication au cours desquels le guide va, tout en continuant de s'intégrer de façon crédible dans l'environnement, présenter des informations au visiteur sur un sujet particulier. On peut également distinguer une autre catégorie de comportement, utilisés au sein de notre application par exemple : les comportements permettant d'interagir avec le visiteur mais ne

délivrant pas d'informations sur l'environnement. C'est le cas dans notre application lors des phases d'introduction et de conclusion d'une session d'explication. Durant ces phases, le guide virtuel se présente et décrit le déroulement de la visite ou fait comprendre au visiteur que la visite est terminée. Ce sont ici des informations sur la visite elle-même et non plus sur ce que le visiteur est venu observer et leur présentation nécessite généralement la mise en place de scénarios statiques du fait que l'état de l'environnement n'intervient pas dans leur déroulement.

3.6 Conclusion

Nous avons présenté dans ce chapitre l'architecture de prise de décision du guide virtuel autonome que nous proposons. Ce guide a pour objectif d'expliquer au visiteur un environnement réel dynamique et peu prévisible dans lequel il est lui-même immergé de façon crédible. Les explications qu'il fournit le sont au travers d'un ensemble de modalités visuelles et sonores.

La définition des explications se fait de façon modulaire à l'aide d'éléments atomiques simples organisés les uns par rapport aux autres au sein d'un graphe. Ce graphe décrit les contraintes et relations entre les différents éléments du discours du guide virtuel. La notion de sujet (processus d'explication) permet de définir l'exécution d'une telle explication et de lui adjoindre des informations sémantiques sous la forme d'un ensemble de mots clés. Le choix de l'explication devant être élaborée face au visiteur est effectué à l'aide d'un ensemble d'experts spécialisés suivant différents points de vue et capable de déterminer la pertinence d'un sujet spécifique par rapport au contexte de la visite. Le passage d'une explication à une autre est rendu possible grâce à l'utilisation de transition dont la validité est définie par rapport à ce même contexte.

La gestion du comportement est effectuée à l'aide d'un automate hiérarchique séquentiel composé d'états élémentaires et d'états dits composites. Ces derniers sont constitués d'un ensemble de sous-automates pouvant être combinés de façon séquentielle ou concurrente suivant les besoins de l'application.

Nous avons défini le concept de visite guidée au travers de l'évolution conjointe du comportement et des explications du guide virtuel. La gestion de l'aspect dynamique du système qu'implique le type d'environnement réel auquel nous nous intéressons est rendue possible au travers de différents mécanismes. Les explications tout d'abord : leur définition à partir d'éléments atomiques simples et la notion de processus d'explication nous permet de gérer le contenu de la visite comme un ensemble d'explications spécialisées concurrentes qui sont en compétition afin d'être présentées au visiteur. Chaque explication pouvant être interrompue entre les différents éléments atomiques qui la constitue, plusieurs explications peuvent se voir exécutées en parallèle, une seule pouvant être effectivement présentée à un instant donné. Le comportement ensuite : la notion d'interruptibilité au niveau des états élémentaires de l'automate hiérarchique séquentiel qui détermine le comportement du guide permet un contrôle précis sur les actions du guide et le moment où celles-ci peuvent être stoppées.

Les différents points que nous avons présentés jusqu'ici se sont donc concentrés sur le discours et les processus de gestion de comportement et de prise de décision du guide virtuel

que nous proposons. Il reste néanmoins un aspect incontournable à traiter : le problème de la perception visuelle de l'environnement à expliquer. Lors de la description du contexte de la visite guidée, nous avons fait l'hypothèse de l'existence d'une représentation (éventuellement incomplète) cohérente de l'environnement réel à décrire. La partie suivante de ce manuscrit s'attache à détailler les problèmes liés à cette perception de l'environnement réel et à la construction d'une représentation exploitable par le processus de décision que nous venons de présenter.

Partie III

— Percevoir le réel —

Chapitre 4

Représenter le réel pour le décrire

4.1 Introduction

Dans ce chapitre nous nous intéressons aux travaux existant dans le cadre de la vision artificielle et plus particulièrement à la notion de perception active. Nous commençons par offrir un aperçu des différentes théories de la perception visuelle avant de nous intéresser aux approches de la vision artificielle. Nous présentons ensuite une vue d'ensemble du domaine de la vision artificielle avant de nous attarder sur le problème de l'attention visuelle.

4.2 Contexte et contraintes

Le rôle de notre guide virtuel est d'expliquer à des visiteurs l'environnement réel au sein duquel il est immergé. Pour cela il doit être en mesure de connaître l'état de cet environnement et donc de le percevoir. Ce problème de la perception visuelle d'un environnement réel est complexe.

Dans les problèmes classiques de la réalité augmentée, l'environnement réel est généralement fixe ; il suffit de connaître l'état de l'utilisateur (orientation et/ou position) afin de pouvoir le placer au sein d'une scène virtuelle. C'est cette scène virtuelle statique qui est ensuite superposée au monde réel à l'aide de différents systèmes de visualisation selon le point de vue de l'utilisateur. Le cas que nous abordons est plus complexe au sens où, ici, ce n'est pas le point de vue du visiteur que l'on cherche à reconstruire mais l'environnement du guide virtuel. Notre système de perception doit permettre de connaître l'état et l'évolution de l'environnement réel afin de mettre à jour la scène dans laquelle le guide virtuel évolue.

Nous considérons que la perception de l'environnement réel par le guide virtuel s'effectue à l'aide de capteurs passifs (caméras). Percevoir l'environnement doit donc passer par la mise en œuvre de traitements sur les données provenant de ces capteurs (images) afin d'en extraire des informations utiles et exploitables par le processus de prise de décision. Pour cela nous devons avoir recours à des techniques propres au domaine de la vision par ordinateur (cf. 4.5.1). Le système de perception visuelle du guide virtuel doit être en mesure de détecter,

suivre et reconnaître les différentes entités présentes dans l’environnement, et ce en respectant un certain nombre de contraintes.

Le guide doit être capable de percevoir l’environnement en temps réel, c’est à dire que le module de perception doit être en mesure d’assurer une certaine cohérence temporelle entre l’état réel de l’environnement et sa représentation. La notion de temps réel est souvent opaque. Dans notre cas, cela correspond à l’objectif suivant : les explications énoncées par le guide sur son environnement doivent être en accord avec l’état de cet environnement au moment où il présente ces explications. Il n’est en effet pas envisageable de décrire une entité qui n’est plus présente au moment de l’explication, ou du moins de la décrire en la supposant présente.

Le dernier aspect non négligeable est ici l’environnement observé. Il est important de noter que nous ne faisons aucune supposition particulière quant aux conditions d’éclairage ou de pose des entités que l’on souhaite percevoir. La seule contrainte que nous fixons ici est l’utilisation d’un point de vue statique par rapport à la scène observée (*i.e. les caméras ne bougent pas*). Ceci, allié à la contrainte du temps réel, implique l’impossibilité de traiter la totalité du champ visuel en permanence et donc la nécessité de mettre en œuvre des mécanismes de prise d’information adaptés.

4.3 Perception visuelle

Le problème de la perception visuelle et plus particulièrement la perception visuelle humaine est le centre d’intérêt de nombreuses recherches depuis l’antiquité jusqu’à nos jours et ce dans plusieurs domaines (biologie, neurobiologie, psychologie, robotique, informatique).

Différents courants se sont affrontés au cours de l’histoire [Harrison 02] donnant naissance à de nombreuses théories quant au fonctionnement de différents aspects de la perception [Shimon 92; Delorme 03]. Nous présentons ici les grandes tendances à travers deux questions centrales qui abordent le problème de la perception selon des points de vue généralement opposés :

- ▷ La perception se fait-elle à partir d’éléments simples qui sont ensuite composés ou est-elle plus globale ?
- ▷ La perception de l’environnement dans son ensemble se fait-elle directement ou indirectement par inférence de perceptions locales ?

Sur cette dernière question nous présentons un point de vue intermédiaire : la notion de perception interactive et de perception en fonction de la tâche. Enfin nous nous intéressons au problème de la mémoire et de la représentation de l’environnement qu’elle implique.

4.3.1 Une vue élémentaire ou globale ?

La théorie structuraliste développée par Titchener dans les années 1900 explique la perception d’une façon générale comme pouvant être synthétisée à partir d’un ensemble de sensations.

Ces théories ont été peu appliquées dans le cadre de la perception visuelle en informatique. Par opposition, la théorie des Gestalt (théorie de la forme ou de la configuration) [Gordon 89] décrit la perception comme étant la distinction de configurations particulières dans le champ visuel. L'idée directrice de cette théorie est l'émergence et le fait que le tout est différent de la somme de ses parties. Ainsi le tout est généralement perçu avant les parties qui elle-même peuvent prendre un sens différent suivant le tout dans lequel elles sont incluses. La structuration des formes dans cette théorie se fait selon certaines lois dites « naturelles » qui s'imposent au sujet. Parmi ces lois on trouve :

- ▷ La loi de la bonne forme : un ensemble de parties informes tendent à être perçues comme une forme simple, symétrique et stable.
- ▷ La loi de bonne continuité : des points rapprochés tendent à représenter une continuité.
- ▷ La loi de proximité : nous regroupons avant tout les points proches les uns des autres.
- ▷ La loi de similitude : nous regroupons ensuite les points ayant des similarités (même couleur par exemple).
- ▷ La loi de destin commun : des parties en mouvement ayant la même trajectoire sont perçues comme faisant partie de la même forme.
- ▷ La loi de clôture : nous préférons les formes fermées aux formes ouvertes.

Dans cette théorie, la perception et la cognition se font sur le tout et non sur les parties. Des expérimentations ont permis de mettre en évidence les phénomènes perceptifs décrits par cette théorie en se basant sur des mécanismes plus élémentaires [Pomerantz 89].

Néanmoins, la critique majeure de cette théorie vient de la difficulté de définir objectivement ces formes et du fait que c'est une théorie uniquement descriptive qui n'explique pas comment se forment ces configurations.

4.3.2 Perception directe ou inférence ?

L'approche écologique de la perception [Gibson 79] définit la perception comme un ensemble animal/environnement plutôt que comme le seul fait d'une interprétation du cerveau. Il introduit la notion d'*affordance* qui est liée aux capacités de l'être vivant dans son environnement. Ainsi lorsqu'il perçoit son environnement un animal perçoit ces *affordances*, des possibilités d'action dans cet environnement, qui sont inévitablement liées à ses capacités. Ici on perçoit dans un but précis et la prise d'informations reflète cet objectif. De plus cette perception est directe, aucun traitement n'est nécessaire pour percevoir les affordances dans l'environnement (le flux optique est un exemple d'une telle perception utilisé par Gibson). Cette théorie a trouvé des applications dans le domaine informatique et notamment pour la perception des acteurs virtuels au sein d'environnement virtuels. Ici l'environnement est vu comme une base de données dans laquelle on va sélectionner les informations qui nous intéressent [Bordeux 99; Herviou 06].

Une approche opposée est celle proposée par Helmotz et dont la théorie de Marr [Marr 82] est en quelque sorte la continuité. Helmotz, partant du constat de la faible qualité des informations perçue au niveau de l’œil, voit la perception comme une inférence sur les stimuli visuels, inférence qui permet de décider de la validité d’hypothèses sur notre environnement. La formation de ces hypothèses implique ici un traitement symbolique sur les stimuli visuels, traitement qui se fait de façon inconsciente. Marr propose une méthodologie pour la réalisation de systèmes de perception basés sur cette idée. Il propose de découper le phénomène de perception en plusieurs niveaux :

- ▷ Le niveau computationnel qui traite du problème de vision lui même, de ce que l’on cherche à accomplir au travers de la perception.
- ▷ Le niveau algorithmique qui traite des stratégies pouvant être utilisées.
- ▷ Le niveau implémentation qui décrit la façon dont cela s’effectue réellement au niveau neuronal.

Il décrit la vision comme une succession de trois étapes (*primal sketch* ; *2,5D sketch* ; *3D model*) pendant lesquelles le système visuel part des stimuli (2D) afin de construire une représentation tridimensionnelle continue (et totale) de l’environnement indépendante du point de vue (centrée objet). Cette approche de la perception visuelle peut être considérée comme l’approche classique de la vision artificielle (cf. 4.4).

4.3.3 Percevoir en fonction de la tâche

Dans [Churchland 94], l’approche computationnelle de Marr est critiquée sous le terme de *vision pure* en faveur d’une notion de *perception interactive*. Dans cette théorie, il n’y a pas reconstruction d’une représentation complète de l’environnement mais une prise d’information appropriée en fonction des intérêts ou des buts. Ainsi, une représentation partielle (incomplète) mais suffisante pour guider la focalisation de l’attention est maintenue en mémoire. Cette représentation est fortement liée aux actions moteur de l’animal, que ce soit les saccades de l’œil ou le fait de se déplacer. Le système visuel y est également intimement lié aux autres sens tels que l’ouïe. Une telle théorie s’appuie sur des faits neurobiologiques et psychophysiques et met en avant l’importance des interactions au sein et à l’extérieur du système visuel. Elle montre aussi l’aspect central de l’attention dans le maintien de la représentation partielle de l’environnement. Cette idée d’une perception intimement liée à la tâche que l’on effectue et aux mécanismes attentionnels s’est vue soutenue plus récemment au cours d’expérimentations chez l’Homme [Triesch 03].

Dans le domaine de l’intelligence artificielle, on retrouve également cette idée d’une perception interactive liée aux actions. [Wheeler 94] constate notamment que la perception (pas nécessairement visuelle) part d’une évolution pour une tâche particulière et est de ce fait liée aux actions pour lesquelles elle s’est développée.

Dans [Brill 94] une idée similaire est soutenue. L’auteur insiste sur la nécessité de mécanismes attentionnels afin de palier aux ressources computationnelles limitées :

“We only want to represent those aspects of the environment that are task-relevant [...] moreover, since computational power is bounded [...] focus of attention strategies are needed to concentrate the resources where they are most likely to produce results” – [Brill 94]

4.3.4 Mémoire et représentation

Un dernier aspect important de la perception est la mémoire, celle-ci est nécessaire pour accomplir certaines tâches complexes comme se souvenir de la position d’objets au cours du temps par exemple [Brill 94]. Atkinson et Shiffrin [Atkinson 68] proposent une décomposition de celle-ci selon trois niveaux déterminés par le temps de rétention des informations (figure 4.1) :

- ▷ La mémoire sensorielle qui est caractérisée par un temps de rétention des informations très court (de l’ordre d’une centaine de millisecondes). Des études ont montré que cette mémoire est capable de stocker une douzaine d’objets [Sperling 60] ;
- ▷ La mémoire à court terme ou mémoire de travail d’une capacité d’environ 7 ± 2 objets [Miller 56] (bien que récemment ce chiffre ait été revu à la baisse (4–5 objets) [Cowan 01]). Le temps de rétention des informations dans cette mémoire est de l’ordre de quelques secondes ;
- ▷ La mémoire à long terme d’une capacité *a priori* non définie et capable de stocker des informations sur une période de temps illimitée (à l’échelle de la vie humaine). Cette mémoire amène de l’abstraction et serait basée sur la sémantique des informations stockées.

Les auteurs décrivent le passage des informations d’une mémoire à l’autre comme se faisant :

- ▷ Au moyen de l’attention pour le passage de la mémoire sensorielle à la mémoire à court terme.
- ▷ Par renforcement (la même information qui revient plusieurs fois) pour le passage des informations de la mémoire à court terme vers la mémoire à long terme et remémoration (accès volontaire à une information existante) pour le passage d’informations de la mémoire à long terme vers la mémoire à court terme.

Dans la théorie de [Craik 72] cette séparation en différents niveaux est mise de côté au profit d’une vision plus continue. Ici le niveau de mémorisation des informations dépend du niveau des traitements qui leurs sont appliqué, une analyse plus complète permettant une rétention plus longue des informations.

Plus tard, Baddeley et Hitch [Baddeley 74] ont montré que la mémoire à court terme servait de zone de travail dans laquelle nous transformons les informations. Il ont donc proposé un modèle plus fin de cette mémoire à court terme (figure 4.2) composé des éléments suivants :

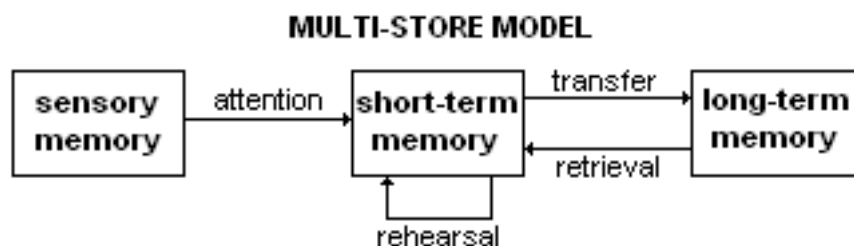


FIG. 4.1 – Modèle de mémoire proposé par [Atkinson 68]

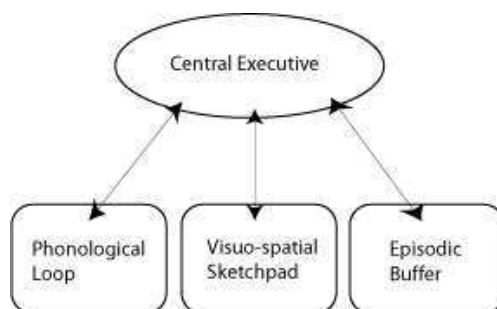


FIG. 4.2 – La mémoire de travail proposée par [Baddeley 74]

- ▷ Le coordinateur central qui sert de mécanisme attentionnel et contrôle les deux autres mécanismes.
- ▷ La boucle phonologique pour la mémoire des sons.
- ▷ Le calepin visuo-spatial pour la mémoire visuelle et spatiale.

Tous ces modèles de mémoire supposent un codage des informations sous une forme ou sous une autre (codage verbal et visuel par exemple pour le modèle proposé dans [Baddeley 74]).

Dans l'approche représentationnelle de Marr [Marr 82], le système visuel a pour rôle de reconstituer une représentation tridimensionnelle complète de la scène perçue. Au contraire l'approche proposée par Brooks [Brooks 91a] nie le besoin de toute forme de représentation alors que l'approche écologique de Gibson [Gibson 79] considère l'environnement comme étant sa propre représentation.

Dans [Wheeler 94] l'environnement en tant que représentation est remis en cause et plutôt que de se demander si une représentation (un état interne des perceptions) est nécessaire, l'auteur propose de se poser la question : « quelle représentation pour quelles actions ? ». Dans [Brill 94] l'absence de représentation est également remise en cause et notamment du fait de la nécessité de la mémoire pour certaines tâches (exemple d'un animal qui doit se souvenir de la position d'un prédateur afin de l'éviter).

L'approche de Rensink [Rensink 00] quant à elle suppose l'existence d'une mémoire partielle dans laquelle les objets seraient formés par le processus attentionnel permettant de capturer leur structure (cf. 4.6.4).

Un dernier aspect de la mémoire est celui de sa mise à jour. Dans [Fabiani 96] ce problème est abordé dans le cadre d'un environnement dynamique complexe (au sens où bien que ses règles d'évolution soient connues il n'est pas possible de considérer toutes les éventualités à un instant donné) et de contraintes de calcul fortes. Pour faire face à cette complexité de l'environnement le système autonome doit être en mesure de déterminer quand récupérer des informations à l'aide de ses capacités de perception et doit donc pour cela posséder un moyen de déterminer son ignorance par rapport à la réalité du monde. Les auteurs présentent un ensemble de techniques de modélisation de l'ignorance ou réciproquement de la confiance dans les connaissances courantes de l'agent et notamment les mécanismes d'érosion.

4.4 Approches de la vision artificielle

Les systèmes de vision s'inspirent généralement du paradigme de Marr [Marr 82] et abordent la vision comme une succession de trois étapes (segmentation, reconstruction et reconnaissance) afin de construire une représentation tridimensionnelle de l'environnement. Dans la description de sa méthodologie, Marr va même jusqu'à nier l'utilité de toute connaissance *a priori* dans le processus de vision humain.

Ces systèmes reposent donc sur une approche purement computationnelle basée sur l'utilisation de capteurs de façon passive au sens où leurs paramètres (position, zoom, ...) ainsi que les traitements effectués sont les mêmes quel que soit le problème abordé. Le problème d'une telle approche réside dans le fait qu'une solution unique à tous les problèmes de vision n'existe pas. De plus les études du système visuel humain tendent à montrer un rôle fondamental des connaissances dans le processus de perception visuelle notamment au travers des mécanismes attentionnels (cf. section 4.6).

Pour contrer cette limitation, un certain nombre de chercheurs ont rapidement proposé d'autres approches du problème de la vision en s'inspirant notamment de la vision humaine. Nous présentons ici quelques uns de ces courants regroupés sous le terme général de *perception active*. Celle-ci consiste en l'utilisation de capteurs passifs de façon active. Cela peut passer par le contrôle du matériel (mouvements de caméra) et/ou le contrôle du logiciel (choix/paramétrage des traitements à effectuer, mécanismes attentionnels) [Swain 93].

4.4.1 Vision active

Aloimonos [Aloimonos 88] fut l'un des premiers à explorer ce concept. Il considère les problèmes fondamentaux de la vision comme étant mal posés dans le cadre du paradigme de Marr. Il reformule ces problèmes en considérant les contraintes supplémentaires issues du mouvement des capteurs. Il s'est intéressé à plusieurs problèmes particuliers et a mis en évidence leur simplification grâce aux nouvelles contraintes. Certains auteurs considèrent cette

approche comme de la *vision dynamique* du fait d’une formulation uniquement mathématique, de l’absence de vérifications expérimentales et du fait que les capteurs ne sont que peu pris en compte [Marchand 96].

4.4.2 Perception active

La notion de perception active introduite par Bajcsy [Bajcsy 88] se veut légèrement différente de la notion de vision active au sens où les auteurs s’intéressent ici à l’élaboration de stratégies de perception permettant d’améliorer la connaissance de l’environnement. Ainsi la perception active consiste en une boucle permanente entre l’acquisition d’informations à partir des capteurs à l’aide d’algorithmes et le contrôle de ces algorithmes à l’aide des informations perçues.

“The fundamental ingredient of the Active Vision theory is the mechanism for decision making in general, how to choose data information with respect to the perceptual goal. This involves both the method of estimation and the choices of general strategy for control.” – Ruzena Bajcsy [Bajcsy 88]

4.4.3 Vision animée

Ballard [Ballard 91] s’intéresse au problème plus spécifique du contrôle de la vision par rapport à la tâche et notamment à la coordination entre le comportement et le regard. L’une des particularités de cette approche est sa restriction à une vue binoculaire. Les auteurs s’intéressent ici à la simplification des problèmes de vision du fait de contraintes supplémentaires connues dues au mouvement des capteurs. Ainsi les différents traitements se font ici de façon relative par rapport au point de fixation (*fixation frame*). Cette approche de la vision active est fortement inspirée de systèmes visuels réels et du constat de la nécessité des mouvements de l’œil notamment dans la perception visuelle humaine. Un point notable de cette approche est l’exclusion de toute représentation explicite de l’environnement.

4.4.4 Vision intentionnelle

[Aloimonos 90] propose quant à lui la notion de *vision intentionnelle* qui cherche à étudier les problèmes de la vision par ordinateur dans le cadre de problèmes spécifiques, prenant en compte l’objectif du système de vision dans son développement. Les auteurs s’éloignent ainsi de la quête classique de la vision par ordinateur qui consisterait en la création d’un système générique unique capable de résoudre n’importe quel type de problème. Ils voient au contraire les futurs systèmes de vision comme des ensembles composites d’algorithmes spécialisés associés à des mécanisme de prise de décision permettant de sélectionner ces algorithmes en fonction d’un objectif bien défini.

“[...]Consider a general vision system of the futur as we envision it [...]. That system will consist of a large number of modules, each of which will be devoted to recovering a property of the world from a series of images. [...] And all these modules will communicate and cooperate in building an accurate description of the environment [...]. [...] this general system will have many more high-level modules which, using the results of the other modules, will perform planning and reasoning. [...]” – John Aloimonos [Aloimonos 90]

La vision intentionnelle se différencie de la perception active au sens où l'on s'intéresse ici à l'organisation de l'exécution de stratégies prédéfinies plutôt qu'au paramétrage d'algorithmes de traitement d'image.

4.5 Vision artificielle

4.5.1 Vision par ordinateur

La vision par ordinateur est un domaine large et qui a longtemps cherché à se définir correctement. D'une façon générale, on peut considérer comme faisant partie du domaine de la *vision par ordinateur* l'ensemble des sciences et techniques permettant aux machines de « voir »¹.

La définition d'un but pour le domaine de la vision par ordinateur dépend fortement des intérêts des différents protagonistes qui y participent. Ainsi certains cherchent à recréer le système de vision humain tandis que d'autre cherchent à créer un système de vision plus générique². D'autres enfin cherchent à mettre en place des algorithmes de traitement ou d'apprentissage pour des problèmes bien identifiés et considèrent ceci comme l'objectif final de la vision par ordinateur [Chetverikov 95].

La vision par ordinateur repose à son plus bas niveau sur la mise en œuvre de traitements d'images. Ces traitements reposent généralement sur l'utilisation de techniques propres au domaine du traitement du signal en considérant l'image comme un signal bidimensionnel. Le résultat de ces traitements peut consister en une nouvelle image modifiée (application de filtres) ou l'extraction de caractéristiques (*features*) particulières telles que des contours. La plupart des traitements sont aujourd'hui connus et éventuellement adaptés à des applications particulières, nous ne nous étendrons donc pas sur ce problème. Le lecteur pourra se référer à [Ballard 82] et [Jähne 99] pour plus d'informations.

Basé sur ces traitements bas niveau, différents problèmes peuvent être identifiés parmi lesquels :

¹ Ceci inclus bien évidemment l'utilisation de caméras vidéos classiques mais également l'utilisation de capteurs moins ordinaires tels que des radars ou des caméras infrarouge.

² Partant généralement du constat que bien que le système visuel humain soit extrêmement performant il est des tâches pour lesquelles la machine est malgré tout plus efficace.

- ▷ la segmentation qui consiste en la détection de zones « cohérentes » au sein d’une image. Ces zones correspondent généralement à un objet unique, la notion d’objet variant fortement suivant le point de vue.
- ▷ le suivi qui concerne le mouvement ; ici on cherche à mettre en correspondance des objets d’un instant à un autre. Deux approches principales existent dans la littérature : les approches basées modèle et les approches basées image. La première approche consiste en la recherche d’un modèle connu dans l’image. La seconde repose sur la mise en correspondance de primitives caractéristiques (coins par exemple) d’une image sur l’autre au sein d’une séquence vidéo. Le lecteur pourra se référer à [Allen 01] pour une vue d’ensemble des problèmes de suivi et des techniques généralement utilisées pour les résoudre.
- ▷ la reconnaissance des objets qui consiste en leur classification suivant différents types. La résolution de ce problème repose classiquement sur l’utilisation de classifieurs permettant de catégoriser un vecteur de caractéristiques dans un espace déterminé selon un ensemble fini de classes distinctes [Zheng 05b]. Une autre approche consiste en des mécanismes de production/vérification d’hypothèses.

Dans ce manuscrit nous ne cherchons pas à proposer de nouveaux algorithmes ou méthodes pour le traitement d’image ou les problèmes classiques de la vision par ordinateur. Au contraire, nous cherchons à trouver des mécanismes permettant de choisir et de paramétrer des algorithmes connus afin d’effectuer les traitements appropriés en fonction de la tâche courante du processus de prise de décision du guide virtuel. Le lecteur pourra se référer à [Rosenfeld 99] pour une bibliographie relativement complète sur les différents problèmes du traitement d’image et de la vision par ordinateur. Notre intérêt se porte ici sur les mécanismes de contrôle au sein des systèmes de vision et leur interaction avec les mécanismes de prise de décision.

4.5.2 Les systèmes de vision : du traitement à l’interprétation

Au delà de ces problèmes spécifiques reposant sur la mise en œuvre d’algorithmes de traitement, le domaine de la vision par ordinateur comporte un aspect intelligence artificielle au travers des problèmes liés à la compréhension et l’interprétation d’images. La compréhension d’images consiste en la transformation de stimuli visuels en une description de l’environnement exploitable par un processus de prise de décision [Tsotsos 92]. Ce problème concerne l’aspect haut niveau de la vision et se distingue de la simple reconnaissance d’objet au sens où la description de l’environnement recherchée est plus générale, elle ne consiste pas simplement en un étiquetage des stimuli visuels mais en la création de connaissances à partir :

- ▷ des informations qui sont extraites des images ;
- ▷ de connaissances *a priori* sur le domaine d’application ;
- ▷ de connaissances sur les traitements possibles/disponibles.

Cette description de l'environnement peut se présenter sous de multiples formes allant de simples valeurs booléennes associées aux points d'une image à une reconstruction tridimensionnelle complète de l'environnement ou encore à la reconnaissance d'interactions entre les différentes entités perçues.

L'idée principale des systèmes de vision est, par opposition à la vision bas-niveau, le fait qu'une approche ascendante simple ne suffit pas et conduit à une explosion combinatoire [Tsotsos 01]. La prise en compte de connaissances permet de limiter cette explosion en offrant la possibilité de rechercher une solution donnée (un modèle) contraignant le nombre de solutions à envisager pour l'interprétation des stimuli visuels.

Plus récemment la notion de *vision cognitive* a été introduite pour désigner des systèmes de vision robustes capables d'adaptation face à des conditions non prévues lors de leur conception [Vernon 04]. La vision cognitive repose sur l'idée d'un ajout de capacité cognitive et de connaissances aux systèmes de vision classiques afin de leur permettre de s'adapter et d'apprendre. Elle rejoint de ce fait ce que nous désignons ici sous le terme de *systèmes de vision*³.

Dans la suite de cette section nous nous intéressons à trois aspects de ces systèmes :

- ▷ la représentation des informations ;
- ▷ les architectures logicielles et les mécanismes de coopération ;
- ▷ les mécanismes d'adaptation et de contrôle.

Notre présentation de ces différents aspects est loin d'être exhaustive et nous cherchons avant tout à mettre en avant la variété des approches existantes. Le lecteur pourra se référer à [Tsotsos 92; Garbay 00; Boucher 99; Rao 88] pour des informations plus détaillées sur ces aspects dans les systèmes de vision et à [Draper 89; Hanson 78; Matsuyama 89; Thonnat 93; Moisan 95; Howarth 92; Sandalky 95; Brooks 81; Paulus 99; Crowley 94] pour des systèmes de vision particuliers.

4.5.3 Représentation des informations

Les systèmes de vision doivent manipuler de nombreuses informations afin de construire une description de la scène observée à partir des images fournies par les capteurs. Deux types d'informations sont ici à distinguer : les informations extraites des images et les connaissances *a priori*, connaissances exogènes fournies lors de la conception du système de vision [Garbay 00; Tsotsos 92]. Ces informations peuvent concerner des domaines variés tels que :

³ Cette dénomination est un abus de langage que nous nous permettons étant donné notre intérêt affiché pour le problème du contrôle des traitements et algorithmes en vision par ordinateur. Le terme de *système de vision* désigne plus généralement l'application de techniques de vision par ordinateur à la résolution d'un problème particulier

- ▷ la scène observée, le type d’environnement observé et les objets susceptibles de s’y trouver ;
- ▷ les conditions d’acquisition : paramètres de position, orientation, zoom des caméra ou résolution des images ;
- ▷ les opérations disponibles pour extraire des informations, leurs paramètres ainsi que les valeurs possibles pour ces paramètres ;
- ▷ la tâche à effectuer, le problème que doit résoudre le système de vision.

La modélisation des connaissances *a priori* dépasse le cadre de notre étude, on peut néanmoins distinguer deux formes pour ces connaissances dans les systèmes de vision :

- ▷ les connaissances implicites généralement disséminées au sein même du système ;
- ▷ les connaissance explicites pouvant être modélisées de différentes manières.

Nous nous intéressons ici à la façon de représenter les informations extraites de l’image. Celle-ci varie fortement d’un système à l’autre et dépend généralement de l’application envisagée. Une représentation idéale n’étant d’ailleurs pas forcément souhaitable [Edelman 89]. Nous ne cherchons pas ici à faire une revue exhaustive des différentes représentations existantes, celles-ci étant trop nombreuses. Néanmoins il est possible de mettre en avant certaines caractéristiques des formes de représentation utilisées. Le problème de la représentation des informations est abordé plus en détail dans [Garbay 00; Boucher 99]. [Tsotsos 92] présente une liste détaillé de plusieurs systèmes de vision en précisant la façon dont ceux-ci représentent les informations.

Tout d’abord ces informations peuvent être distinguées suivant le référentiel qu’elles utilisent. Nous en distinguons ici trois types principaux :

- ▷ les représentations centrées objet dans lesquelles les informations sont associées aux objets perçus [Brooks 81] ;
- ▷ les représentations centrées vue dans lesquelles les informations sont exprimées par rapport au point de vue duquel elles ont été extraites [Soyer 00] ;
- ▷ les représentations déictiques au sein desquelles les informations/objets sont représentés par rapport au comportement courant de l’agent qui effectue la perception [Agre 95; Chapman 90].

Les représentations centrées objet permettent l’indépendance des informations par rapport au point de vue et facilitent la mise en commun des informations lorsque celles-ci proviennent de plusieurs capteurs. Les représentations centrées sur la vue lors de l’acquisition permettent quant à elle d’ancrer les informations dans l’espace image plus facilement offrant ainsi des connaissances supplémentaires pour les traitements à venir. Les représentations déictiques présentent pour leur part des avantages du point de vue de la prise de décision en offrant une

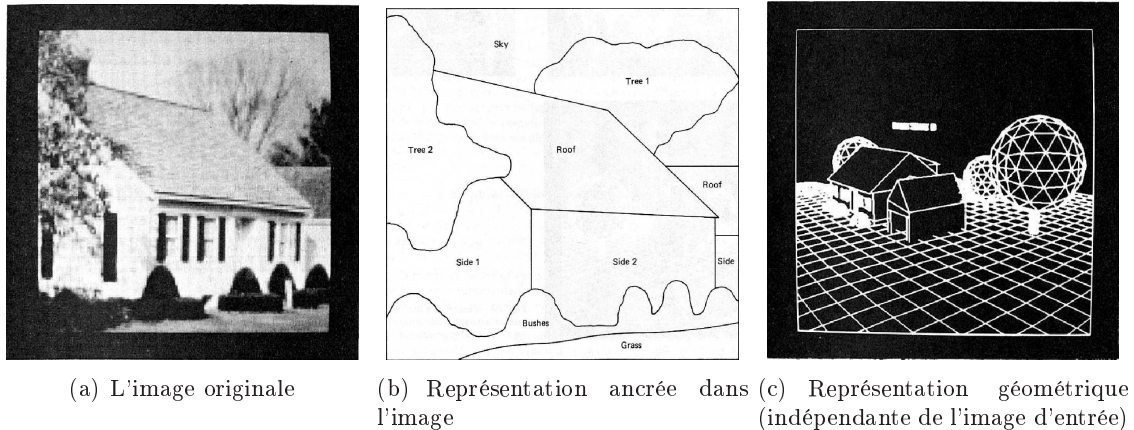


FIG. 4.3 – Différents types de représentation.

indépendance par rapport à l'identité réelle des objets observés, ici on s'intéresse aux objets par rapport à un objectif particulier plutôt qu'à l'objet lui-même.

On peut ensuite distinguer les différentes représentations des informations par rapport à la façon dont celles-ci sont encodées. Ici encore trois approches sont possibles (figure 4.3) :

- ▷ des représentations images qui encodent les informations sous forme de cartes à deux dimensions. C'est le cas des systèmes basés sur les cartes de saillance notamment (cf. 4.6.1) ;
- ▷ des représentations qualitatives qui associent à différentes zones d'une image des informations symboliques ;
- ▷ des représentations symboliques ou géométriques qui stockent les informations indépendamment de l'image dont elles sont extraites.

Ces différents types d'encodage des informations se trouvent souvent combinés au sein des systèmes de vision qui tendent à établir des liens de plus en plus nombreux en organisant ces différentes informations au sein de graphes dont les arcs représentent différents types de contraintes telles que l'appartenance (PART-OF) ou la spécialisation (IS-A) [Tsotsos 92].

4.5.4 Architectures logicielles et coopération

Un système de vision organise la coopération entre différents opérateurs de traitement afin de résoudre un problème spécifique. Ces opérateurs peuvent être organisés de différentes manières et travaillent sur des données et/ou échelles variables. Cette modularité de leur constituants se ressent au niveau des architectures logicielles de ces systèmes. Il est important de noter que ces architectures sont bien souvent implicites au sein des systèmes de vision et dépendent également de la façon dont les informations y sont représentées. Encore une fois nous ne

cherchons en aucun cas ici à proposer une revue exhaustive des différents systèmes de vision existants mais plutôt à présenter les grands types d'architectures utilisées par ces systèmes et les possibilités offertes par celles-ci.

Un type d'architecture très utilisé dans les systèmes de vision est le tableau noir (*blackboard*). L'idée de ces architectures repose sur la métaphore d'un groupe de spécialistes coopérant à la résolution d'un problème et utilisant un tableau noir comme espace de travail commun. Ainsi, les *blackboard* se basent sur une mémoire partagée entre plusieurs sources d'information⁴. Chacune de ces sources d'information possède une expertise propre et tente d'appliquer cette expertise en utilisant les données présentes dans la mémoire venant alors compléter cette mémoire avec de nouvelles informations. Plusieurs modèles de *blackboard* dérivent de cette idée relativement simple suivant des mises en œuvre multiples. Ce type d'architecture centrée sur le partage des informations est particulièrement adapté à la mise en place de systèmes de vision hiérarchiques, les différentes sources pouvant travailler simplement sur les informations déjà présentes en mémoire. [Corkill 91] propose une présentation détaillée de ce type d'architecture. Dans le cadre des systèmes de vision, ce genre d'architecture est utilisée notamment par les systèmes SCHEMA [Draper 89] (scènes d'extérieur ; bâtiments), VISIONS [Hanson 78] ou encore MESSIE-II [Sandalky 95].

Le système *CAVIAR* présente un autre type d'architecture modulaire au sein de laquelle ses concepteurs s'attachent à abstraire le contrôle des opérateurs afin de permettre la comparaison de différentes stratégies de prise de décision pour ce contrôle [Bins 05]. Pour cela leur architecture repose sur un ensemble de modules de traitement sous la direction d'un contrôleur central. Chaque module de traitement fournit une description complète de ses possibilités, des données qu'il attend en entrée, des données qu'il produit ainsi que de ses paramètres (valeurs minimum et maximum, incrément, ...). A chaque pas d'exécution du système (le temps que les données d'entrées soient traitées par chacun des modules) chaque module rapporte au contrôleur des informations sur la qualité et la quantité des résultats obtenus. Le contrôle est ici centralisé, l'ensemble des modules sont assemblés au démarrage en fonction de leurs entrées et sorties et le contrôleur détermine quels modules sont équivalents du point de vue des résultats qu'ils peuvent produire. Lors d'une phase d'apprentissage, les meilleurs modules et paramètres sont sélectionnés par rapport à l'objectif visé. La description des modules ainsi que la communication au sein du système sont assurés à l'aide du langage *CVML* [List 04]. Dans [List 05], cette architecture est comparée au système *Psyclone* dans le cadre d'une application de vision par ordinateur. L'architecture de ce système repose sur la notion de *whiteboard* permettant l'écriture et la lecture d'informations. Chacun de ces *whiteboard* conserve les informations pendant une certaine durée permettant l'accès à des informations passées pour les modules qui le nécessite. Comme pour *CAVIAR* chaque module possède une description de ses propres capacités. L'exécution repose sur la notion de concepts ceux-ci décrivant un état du système à un instant donné. Un ou plusieurs de ces concepts peuvent être assignés à chaque module qui sera exécuté lorsque au moins un des concepts sera vrai (l'état de l'environnement sera vérifié). Une telle approche est similaire à l'architecture employée dans le système VAP [Crowley 94] basé sur un ensemble de modules indépendants (processus UNIX) communiquants au travers d'une interface commune définie par un squelette d'application (SAVA pour Squelette d'Application pour la Vision Active).

⁴ L'utilisation du terme source d'information est ici plus large qu'il n'y paraît. En effet, dans une architecture à tableau noir, ces sources peuvent être à la fois productrices et consommatrices d'informations.

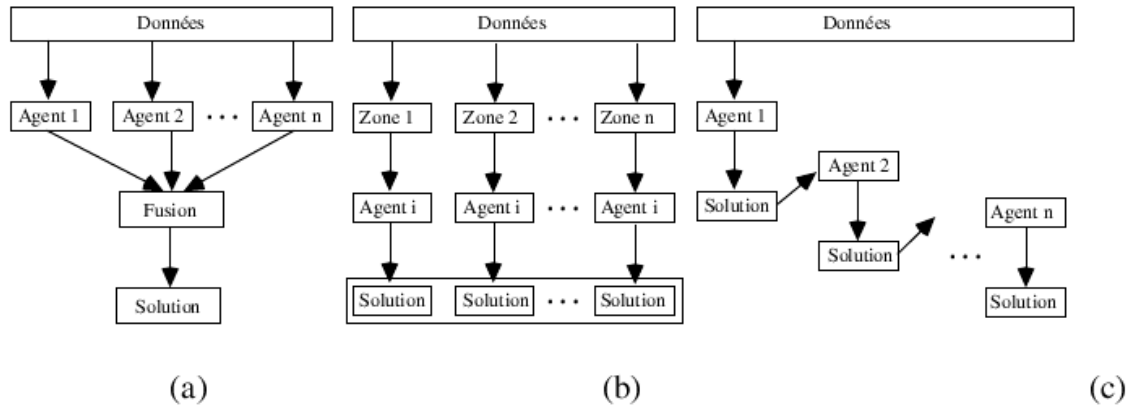


FIG. 4.4 – Différentes formes de coopération entre sources d'informations [Garbay 00]. (a) coopération confrontative; (b) coopération augmentative; (c) coopération intégrative.

Ces architectures sont des exemples de systèmes modulaires tels qu'on les retrouve dans les systèmes de vision. Ces systèmes de vision sont amenés à faire coopérer différentes sources d'information dans le but d'interpréter les images qu'ils traitent. Différentes approches permettent la combinaisons de ces sources d'informations (figure 4.4) :

- ▷ la coopération confrontative qui repose sur l'idée d'exécuter en parallèle plusieurs opérateurs sur les mêmes données. Ce type de coopération peut être abordé de deux points de vue : la mise en compétition d'opérateurs équivalents au sens où ils permettent d'obtenir la même information de façons différentes et la fusion de données nécessaire à l'obtention d'un résultat final ;
- ▷ la coopération augmentative dans laquelle les données sont réparties entres les différentes sources d'information équivalentes. Les sources travaillent ainsi de façon parallèle afin, par exemple, d'augmenter la vitesse de traitement ;
- ▷ la coopération intégrative qui se base sur la décomposition d'une tâche à effectuer en un ensemble de sous tâches attribuées à des sources d'information différentes. Plusieurs types de coopération intégrative sont envisageables et notamment : par décomposition (suivant différents niveaux d'abstraction par exemple), par contrainte (lorsqu'un résultat va contraindre les traitements suivants) ou par dépendance (les systèmes d'appels d'offre par exemple dans lesquelles certaines sources nécessitent les résultats d'autres sources pour mener à bien leurs traitements).

Ces trois formes de coopération peuvent être associées au sein d'un même système de vision, souvent de façon implicite. La coopération confrontative permet de mettre en place des mécanismes de fusion d'informations. Cela peut consister en la fusion de données issues de plusieurs classifieurs pour la reconnaissance par exemple ou encore la prise de décision à partir d'informations conflictuelles. La coopération augmentative permet une redondance des calculs sur des ensembles disjoints de données. Ceci s'avère particulièrement approprié dans le cadre

de la vision où un même objet peut par exemple se trouver en différents endroits d'une image. Enfin la coopération intégrative permet la décomposition d'un problème, sa distribution et son organisation.

4.5.5 Le contrôle

Le problème du contrôle en vision repose sur trois questions d'après [Garbay 00] :

- ▷ Quoi ? – Que doit on chercher, quels objets, quelles propriétés ?
- ▷ Où ? – Quelles zones de la scène observée doit on explorer ?
- ▷ Comment ? – Quels algorithmes, méthodes ou stratégies doit on utiliser ? Avec quels paramètres ?

Le but du contrôle est d'adapter les traitements effectués par le système de vision en fonction d'une tâche particulière. Ceci nécessite de posséder des connaissances sur cette tâche ainsi que sur les capacités du système lui-même au travers des opérateurs à sa disposition. Au-delà de ces connaissances, cela nécessite la mise en œuvre de mécanismes d'adaptation capables de sélectionner des opérateurs et leurs paramètres ainsi que les zones de l'image sur lesquelles ces opérateurs vont être appliqués. Le système de vision doit enfin être capable d'utiliser ces mécanismes au travers de méthodes de coordination.

Mécanismes d'adaptation

On trouve deux types de mécanismes pour l'adaptation dans les systèmes de vision [Garbay 00] :

- ▷ les mécanismes attentionnels (basés sur des connaissances *a priori*) ;
- ▷ les mécanismes de révision/réparation (basés sur une évaluation *a posteriori*).

Les mécanismes attentionnels cherchent à focaliser le système sur des objets, modèles ou actions spécifiques et à ajuster les paramètres des actions sélectionnées. Ces mécanismes peuvent reposer sur différentes approches :

- ▷ les approches basées sur une représentation symbolique des critères de sélection. Parmi ces critères on peut distinguer :
 - ◊ la pertinence qui guide le système vers des choix adéquats en fonction du contexte ;
 - ◊ la facilité qui amène le système à résoudre les problèmes les plus faciles en priorité dans le but de réduire la difficulté des problèmes restants ;
 - ◊ l'aptitude qui guide le système vers les traitements les plus aptes à résoudre le problème considéré ;

- ◇ la vraisemblance qui guide le système vers les hypothèses les plus vraisemblables ;
 - ◇ la complétude qui oriente le système vers le calcul des informations manquantes.
- ▷ les approches basées sur l'estimation et l'optimisation. Ces approches mettent en jeu une formalisation explicite des choix possibles et des gains qui leurs sont associés.

Les mécanismes de révision/réparation reposent quant à eux sur la remise en cause des informations extraites, des opérateurs et paramètres utilisés. Les mécanismes de réparation reposent sur une évaluation des informations provenant des traitements appliqués aux images. A partir de cette évaluation le système décide s'il doit reprendre les traitements et de quelle manière il doit les reprendre. Ainsi un mécanisme de réparation peut déclencher l'exécution d'une segmentation si aucun des objets recherchés n'est détecté. Les mécanismes de révision consistent quant à eux en la remise en cause des informations extraites jusqu'à présent lorsque de nouvelles informations sont disponibles afin de maintenir la cohérence du système.

Coordination et contrôle

Les mécanismes de contrôle et de coordination des systèmes de vision consistent en l'ensemble des procédures, règles et fonctions qui permettent d'orchestrer les traitements en fonction d'un but recherché. Plusieurs type de stratégies de contrôle peuvent être distinguées : les approches délibératives, les approches réactives et les approches optimales [Garbay 00].

Les stratégies délibératives reposent sur la représentation explicite des critères de sélection. Selon cette approche, la résolution du problème de vision se fait au travers du parcours d'un graphe. Ce graphe permet d'associer les représentations des éléments de la scène à différents niveaux d'abstraction (concepts) aux opérations permettant le passage d'un niveau à un autre (tâches). On distingue les mécanismes de contrôle hiérarchiques (ascendants, descendants ou mixte) basés sur les relations de composition/spécialisation des représentations ; des mécanismes de contrôle hétérarchiques faisant appel à d'autres types de contraintes telles que des informations sur les relations spatiales entre objets. Au sein des mécanismes hiérarchiques, on trouve des approches basées sur la planification (dynamique ou non) dont le fonctionnement est de décomposer un problème posé (souvent par un utilisateur humain) en un ensemble de tâches élémentaires (opérateur et ses paramètres) (système OCAPI [Thonnat 93; Moisan 95]). Ces approches abordent le problème du point de vue de la tâche à effectuer et s'opposent aux approches basées sur la construction de concepts qui elles abordent le problème comme l'exploration d'un espace de modèles (systèmes ACRO-NYM [Brooks 81] ou MESSIE-II [Sandalky 95]). Le contrôle hétérarchique quant à lui repose sur une exploration plus souple pouvant se faire dans l'un ou l'autre des espaces (concepts ou modèles) suivant les besoins du système (systèmes SIGMA [Matsuyama 89], SCHEMA [Draper 89] ou VISION [Hanson 78]).

Les stratégies réactives abordent le problème de la vision comme un problème événementiel dans lequel il faut réagir aux stimuli perçus par le système. On y trouve ainsi des approches reposant sur la notion de machine à état permettant de passer d'une stratégie de perception à une autre en fonction d'événements externes ou internes au système ainsi que des approches hybrides qui tentent d'associer des mécanismes de planification aux mécanismes réactifs purs.

Le système VAP [Crowley 94] repose sur une telle approche en mettant en interaction un ensemble de processus réactifs.

Les stratégies optimales font appel à des mécanismes permettant de sélectionner la meilleure stratégie pour résoudre un problème de vision donné. Cette sélection se fait de façon dynamique par réévaluation lors de chaque étape du traitement en se basant sur des critères de coût, d'utilité ou de performance. L'algorithme de sélection à proprement parler peut ensuite reposer sur l'utilisation de différentes méthodes comme une recherche heuristique (système ANIMALS [Paulus 99]) ou des réseaux bayesiens ([Veyret 04b; Veyret 04c] par exemple).

Centralisé ou décentralisé ?

Le contrôle est un problème complexe et les différentes approches peuvent être considérées suivant une multitude de critères. Un de ces critères dépendant fortement des architectures et applications envisagées est celui de la centralisation ou non du contrôle [Boucher 99].

Les approches centralisées reposent sur la réunion des éléments de contrôle au sein d'une même entité. Cette approche est notamment utilisée dans les systèmes SIGMA [Matsuyama 89], MESSIE-II [Sandalky 95] ou par l'outil OCAPI [Thonnat 93; Moisan 95].

Les approches décentralisées par opposition reposent sur la répartition des informations de contrôle au sein d'entités multiples dans le système. Cette approche est présente dans des systèmes tels que VISIONS [Hanson 78] pour lequel le contrôle est réparti entre différents schémas. De même dans le système VAP [Crowley 94] plusieurs processus indépendants coopèrent chacun gardant un contrôle sur les traitements qu'il effectue.

La centralisation facilite la modélisation et le partage des connaissances entre les différents opérateurs de traitement. Elle permet aussi un contrôle plus adapté, celui-ci s'effectuant à partir d'une connaissance globale du système. Les approches décentralisées apportent robustesse et facilité de distribution/parallélisation des traitements. Il est important de noter que les approches décentralisées doivent être tout particulièrement attentives au problème du coût de la communication entre modules de contrôle.

4.5.6 Bilan

Dans cette section nous avons présenté la vision par ordinateur puis les systèmes de vision (au sens de l'interprétation d'images) intégrant des connaissances permettant d'adapter les traitements à effectuer en fonction d'un objectif défini. Nous nous sommes d'abord intéressés à la façon dont ces systèmes représentent les informations extraites des images ainsi qu'aux architectures logicielles modulaires permettant l'intégration de différents opérateurs de traitement au sein d'un système complet de vision artificielle. Puis nous avons présenté les différents aspects du contrôle au sein de ces systèmes au travers des mécanismes d'adaptation et des approches existantes pour la coordination des traitements.

Les systèmes que nous avons présentés se rapportent à la notion de perception active (cf. 4.4.2) au travers du choix et du paramétrage d'opérateurs de traitements. Ces approches

n'impliquent pas forcément de contraintes temps réel fortes telle que notre cadre d'étude l'impose. De plus, bien qu'elles proposent des mécanismes d'adaptation des traitements en fonction d'un objectif défini, celui-ci est généralement statique au sens où le système répond à une question précise qui lui est posée. Dans notre cas, le processus de perception est nécessaire afin de soutenir une prise de décision. Les besoins en terme de perception du point de vue du processus de prise de décision peuvent varier d'un instant à l'autre en fonction de l'état interne ou de l'environnement du système. Ces besoins variables impliquent des changements dans les objectifs devant être remplis par le système de perception.

Pour résumer, les systèmes de vision artificielle que nous venons de présenter tendent à considérer la vision (au sens large de l'interprétation d'images) comme le but du système alors que nous considérons la vision et son contrôle comme un besoin. Cette notion de perception par rapport à un but, une tâche, est très présente dans le système visuel humain et notamment au travers de des mécanismes attentionnels qui y sont à l'œuvre.

4.6 Attention visuelle

L'attention d'une façon générale est le processus cognitif qui permet de sélectionner les informations importantes dans notre perception (auditive, visuelle, ...).

*“Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed, scatterbrained state which in French is called *distracted*, and *Zerstreuung* in German”* – William James [James 90]

Un exemple classique pour décrire le phénomène attentionnel est celui du *cocktail party effect* qui décrit l'effet de l'attention auditive. Lorsqu'on parle avec quelqu'un dans un environnement bruyant tel qu'une réception, notre attention est focalisée sur notre interlocuteur. Malgré tout nous restons attentifs aux bruits ambiants. Ainsi si l'on vient à entendre notre nom au milieu de ces bruits, il est probable que notre attention sera capturée et désengagée de la conversation dans laquelle nous étions impliqués. Le même phénomène peut se retrouver simplement dans la perception visuelle lorsque l'on distingue un objet particulier sur un fond visuellement chargé.

Le phénomène attentionnel n'est bien évidemment pas simplement réductible à son aspect visuel et concerne la perception d'une façon générale. Face à la complexité des traitements visuels et à l'évidente abondance d'informations que nous propose le champ visuel humain, l'attention visuelle représente néanmoins un aspect considérable de la perception et a fait l'objet de nombreuses études et ce dans différents domaines⁵.

⁵ Le lecteur pourra se référer à différentes bibliographies en ligne sur les principaux domaines sur <http://itb.biologie.hu-berlin.de/~wiskott/Bibliographies/Ver-Vis-Att.html>

Il a été prouvé que des problèmes qui nous paraissent simples comme la recherche d'un objet particulier dans notre champ visuel sont irréalisables à l'aide d'une approche purement ascendante en un temps et mémoire raisonnables (le problème est NP-complet) [Tsotsos 89]. D'autres problèmes de la vision sont d'ailleurs également NP-complets [Tsotsos 01]. Dans le cadre des environnements virtuels, [Hill R.W. 99] présente la nécessité d'un mécanisme attentionnel pour la phase de perception d'un pilote virtuel afin de respecter les contraintes temporelles auxquelles il est confronté. C'est face à l'explosion combinatoire du problème de la perception visuelle que l'attention joue un rôle central en permettant la sélection des informations pertinentes ainsi qu'un contrôle à tous les niveaux du processus de perception visuelle. Malgré ce rôle essentiel au sein de la perception visuelle, les mécanismes attentionnels ne sont que peu considérés dans le cadre de la vision artificielle principalement du fait de cadres d'application bien définis (conditions d'éclairage et de pose contrôlés par exemple) [Rothenstein 06].

Nous ne prétendons en aucun cas proposer une revue exhaustive du problème de l'attention visuelle celui-ci étant encore à ce jour mal compris⁶. Néanmoins il nous paraît important de mettre en avant quelques propriétés essentielles de l'attention ainsi que leur impact dans le domaine informatique. Les aspects qui nous semblent ici importants sont :

- ▷ la vision précoce et notamment la notion de carte de saillance ;
- ▷ les traitements séquentiels et la notion de routines visuelles ;
- ▷ l'influence des connaissances et du contexte sur les mécanismes attentionnels ;
- ▷ les liens entre l'attention et la mémoire.

4.6.1 Vision précoce ou pré-attentive

La vision précoce ou vision pré-attentive correspond à une influence ascendante de l'attention visuelle, c'est à dire une influence provenant directement du champ visuel. Dans le système visuel humain, l'œil et plus particulièrement la rétine ont un rôle essentiel dans la perception visuelle [Duchowski 00]. Les photo-récepteurs ainsi que les nombreuses autres cellules qui la composent lui permette de compresser les informations présentes dans le champ visuel sous la forme de différents canaux. Ces différents canaux permettent de coder des informations telles que les orientations, les couleurs ou l'échelle à partir des stimuli visuels.

La capacité de représentation du système perceptif au niveau pré-attentif n'est cependant pas bien définie. Ainsi, [Cavanagh 01] propose la notion de *sprites*, des schémas de mouvement qui peuvent être détectés directement au niveau de la vision précoce. Il semble néanmoins acquis que des formes de représentation plus complexes (telles que la conjonction de différents attributs primitifs) ne soient pas possibles lors des premiers stades de la vision [Arguin 03].

Treisman & Gelade ont proposé un modèle de cette influence ascendante de l'attention visuelle [Treisman 80]. Dans ce modèle, qui comporte de nombreuses similarités avec le

⁶ Le lecteur pourra se référer à [Duchowski 00] pour un aperçu de l'histoire des théories de l'attention visuelle et à [Palmer 99] pour plus de détails sur le problème de l'attention elle-même.

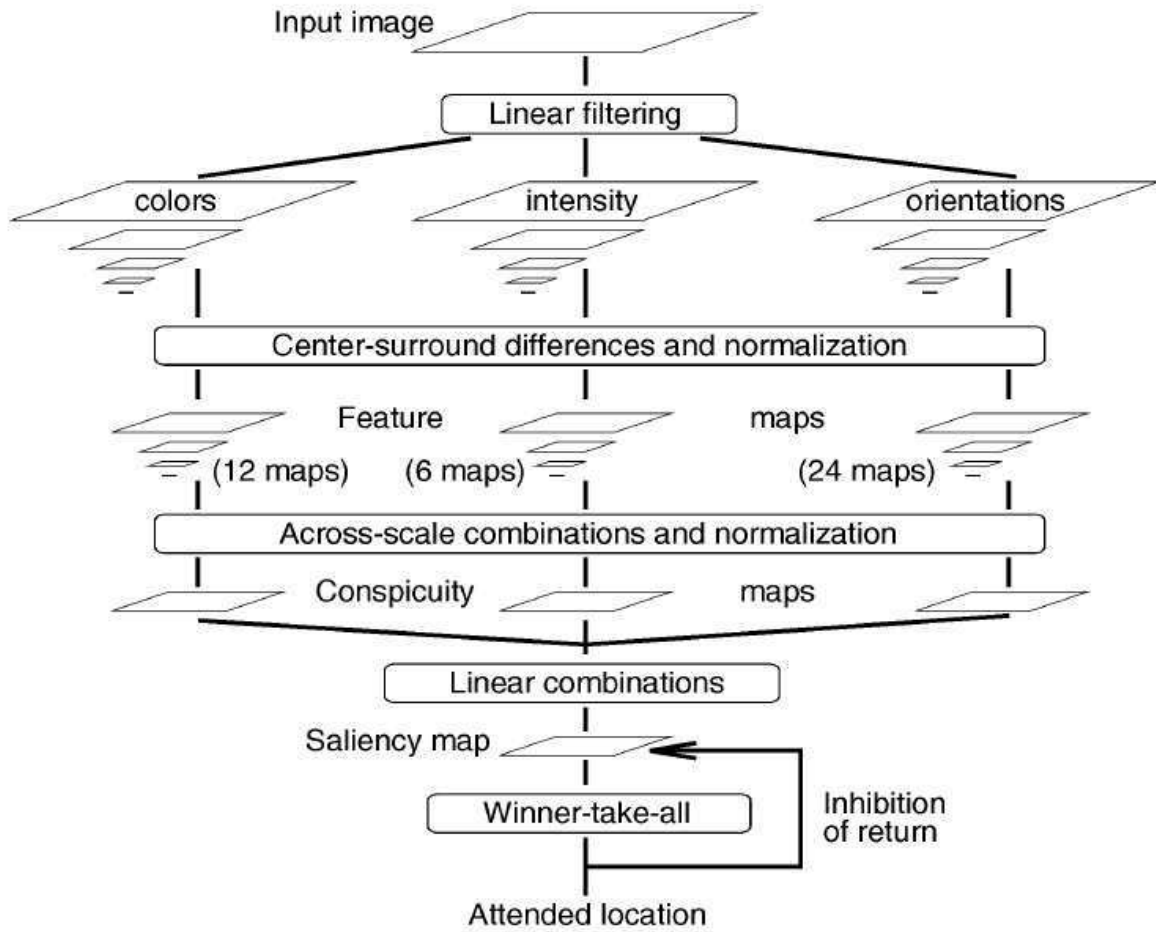


 FIG. 4.5 – Modèle de carte de saillance proposé par Itti & Koch [Itti 98].

guided search de Wolfe [Wolfe 94], les informations du champ visuel sont traitées en parallèle suivant un certain nombre d'attributs parmi lesquels l'intensité lumineuse, des orientations (horizontales, verticales et obliques) ou encore des contrastes de couleurs. Ces informations sont codées au sein de plusieurs cartes d'attributs (*feature maps*) qui sont ensuite intégrées dans une carte de saillance globale qui décrit les zones du champ visuel susceptibles d'être intéressantes (figure 4.6). Dans [Koch 85; Itti 98; Itti 00b] ce modèle est repris et mis en œuvre d'un point de vue informatique (figure 4.5). Dans [Itti 99a], différentes approches d'intégration des cartes d'aspect sont mises en œuvre et comparées.

Ce modèle de focalisation de l'attention visuelle basé uniquement sur les stimuli visuels à été utilisé dans plusieurs travaux et notamment pour guider le regard d'un robot [Chevallier 05], d'un humanoïde virtuel [Courty 03b; Peters 03] ou d'un agent conversationnel animé [Picot 06]. Une implémentation générique du modèle à également été proposée dans [Newcombe 04]. Enfin dans [Itti 05] le modèle est utilisé afin d'évaluer la mise en page de panneaux publicitaires.

Des variations de ce modèle ont également été proposées, notamment dans [Bollmann 97]

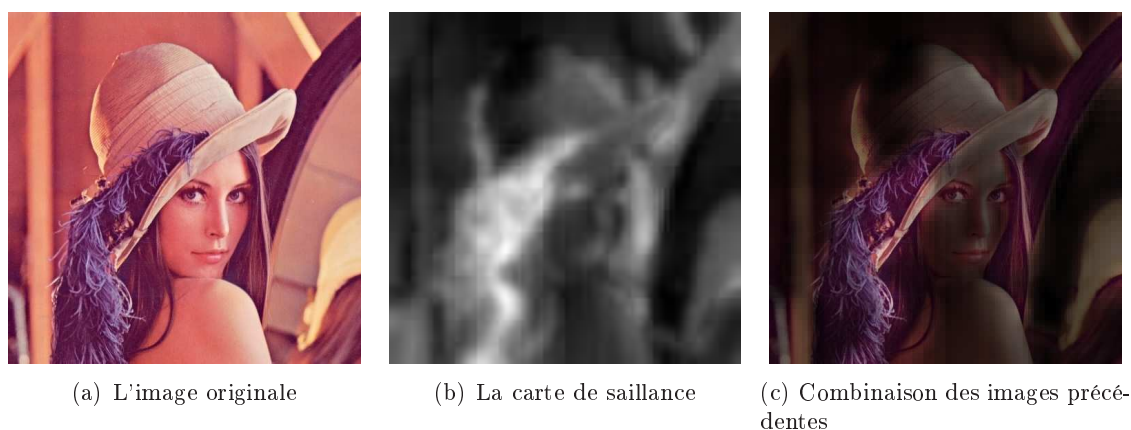


FIG. 4.6 – Exemple de carte de saillance

qui y intègre une influence descendante ainsi qu'un mécanisme d'alerte basé sur le mouvement. Dans [Ouerhani 03] le modèle est étendu afin de prendre en compte la profondeur perçue à l'aide d'une caméra stéréoscopique ainsi que le mouvement à partir du calcul du flux optique. Il est important de noter que ce travail a été validé par rapport à la vision humaine dans le cadre de ces expérimentations. [Tsapatsoulis 03] présente une revue des modèles d'attention basés sur la saillance et les auteurs étendent le modèle afin de prendre en compte le mouvement et des connaissances *a priori* pour la détection de « peau » basées sur des attributs colorimétriques.

Les nombreuses variations de ce modèle d'attention visuelle se différencient généralement par le mécanisme de sélection qu'ils emploient ou la façon dont ils intègrent les différentes cartes d'aspects. Ainsi [Culhane 92] présente le *selective tuning model* ; [Ramström 04a] présente un mécanisme de focalisation basé sur la théorie des jeux afin de n'effectuer qu'une intégration partielle des cartes d'aspect.

Malgré tout ces travaux, certaines questions restent en suspend en ce qui concerne les cartes de saillance [Henderson 03b] :

- ▷ Quelles sont réellement les propriétés à inclure dans la carte de saillance ? Bien que communément acceptées, les propriétés de couleurs, orientation et contraste ne sont qu'un exemple parmi d'autres.
- ▷ Comment combiner l'influence descendante sur la carte de saillance ? Modulation d'une carte existante ou modification du processus de création de cette carte ?
- ▷ Quand doit être calculée cette carte ? Entre chaque fixation⁷ ou après l'exploration complète d'une instance (carte de saillance) ?

En ce qui concerne la première de ces questions, certains travaux ont proposé différentes variations quant à la notion de saillance et aux propriétés à considérer lors de son calcul.

⁷ Une fixation correspond à un mouvement de l'œil, ce mécanisme est décrit plus en détail dans [Duchowski 00] et nous revenons sur ce point dans la section 4.6.2 qui traite des aspects séquentiels de l'attention visuelle.

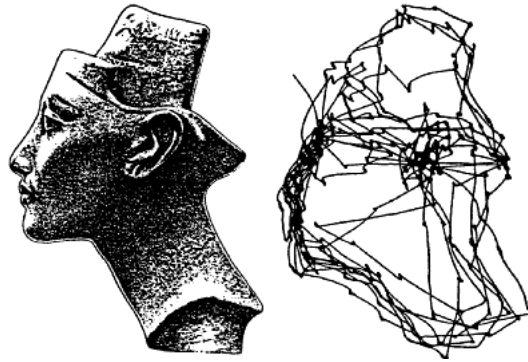


FIG. 4.7 – Parcours de l’œil dans une image [Yarbus 67].

Dans [Olivia 03; Torralba 03; Torralba 05] la saillance est calculée suivant une approche probabiliste permettant de représenter la probabilité de présence d’un objet dans une zone de l’image. Pour [Fritz ; Fritz 04] la saillance se base sur la notion d’entropie décrivant ainsi la capacité d’information des différents points de l’image. Dans [Stentiford 01], une approche évolutionnaire du calcul de saillance est utilisée suivant une notion de dissimilarité d’une zone par rapport à son voisinage.

[Gao 04] propose une notion de saillance intéressante basée sur la discrimination. Contrairement aux approches classiques qui se basent sur des propriétés données (que ce soit les propriétés classiques de couleurs, contrastes, orientation, *etc*, ou les notions de saillance que nous venons de décrire) il considère la définition de la saillance suivante :

“[...] the salient attributes of a given visual concept are the attributes that most distinguish it from all other visual concepts that may be of possible interest [...]”
 – [Gao 04]

Partant de cette définition il présente la saillance comme une notion particulière à chaque objet et fonction des propriétés qui permettent la meilleure discrimination de cet objet par rapport aux autres objets connus. Il est important de noter que cette notion de saillance repose forcément sur des connaissances *a priori* et un nombre de classes d’objets défini à un instant donné.

4.6.2 Traitements séquentiels

Les traitements massivement parallèles qui conduisent à la carte de saillance sont insuffisants pour expliquer les capacités du système visuel humain. Yarbus [Yarbus 67] a montré au travers de son études des mouvements de l’œil que la perception visuelle d’une scène complexe implique un grand nombre de fixations (figure 4.7).

Ullman [Ullman 84] fût le premier à proposer l’idée de traitements séquentiels au travers des *routines visuelles*. Ces routines permettent d’extraire des informations qui ne peuvent

l'être lors des traitements parallèles de la vision pré-attentive comme par exemple les relations spatiales entre différents objets. L'idée d'Ullman est que ces routines sont des primitives de « haut niveau » qui s'appliquent sur les représentations « bas niveau » du système visuel. La particularité de ces routines réside dans leur exécution séquentielle et leur champ d'application limité. Il propose un modèle pour ces routines visuelle au travers de la notion de *visual routine processor* (VRP) et identifie les opérations suivantes comme constituantes primitives des routines :

- ▷ Changement du focus attentionnel ;
- ▷ Marquage d'une zone pour un traitement ultérieur ;
- ▷ Remplissage d'une zone délimitée par un contour ;
- ▷ Suivi de contour ;
- ▷ Marquage d'une position et/ou d'un objet pour utilisation ultérieure.

Combinées entre elles, ces opérations simples permettent d'effectuer des traitements relativement élaborés comme par exemple compter le nombre d'objets qui satisfont une propriété particulière (au sens du codage utilisé dans les cartes pré-attentive) ou encore la reconnaissance d'une forme complexe.

La notion de VRP s'est trouvée implémentée dans différents travaux. Horswill [Horswill 95; Horswill 97a; Horswill 97b] fut l'un des premiers à proposer une implémentation sur des scènes réelles au travers de son système *Jeeves*. Celui-ci permet d'effectuer des requêtes à l'aide d'un langage de logique du premier ordre, requêtes qui sont ensuite transformées sous la forme de routines visuelles afin d'extraire les informations nécessaires des images. Dans [Agre 95; Chapman 90] le programme *Sonja* est capable de jouer à un jeu vidéo uniquement à l'aide d'un système de perception visuelle faisant appel à une représentation déictique et à des routines visuelles. Ici les routines travaillent sur des images artificielles issues du jeu en question et permettent d'extraire des informations telles que *le-bloc-que-va-toucher-le-bloc-que-je-viens-de-pousser*. Dans [Rao 98] un autre gestionnaire de routines visuelles est proposé afin de répondre à des questions d'ordre visuo-spatial telles que « Quel est l'objet pointé par le doigt ? » ou encore « L'objet est-il en train de tomber ? ». L'auteur présente une notion de routine visuelle en accord avec la proposition de Ullman et compare son approche aux VRP existants.

La stratégie de déclenchement des routines visuelles est de nature descendante, c'est à dire en fonction de la tâche à effectuer. Des expérimentations sur le séquençage des tâches dans le cortex visuel primaire viennent soutenir cette théorie des routines visuelles [Roelfsema 03; Deco 02]. Un problème subsiste malgré tout, celui du choix de la routine à utiliser pour accomplir telle ou telle tâche. Ullman désigne ce problème sous le terme de *initial access problem* qui est également appelé *scheduling problem* par d'autres auteurs [Hayhoe 02].

Il est important de noter que nous ne nous intéressons pas ici spécifiquement aux mouvements du regard. Sur ce point [Findlay 01] présente l'attention visuelle du point de vue de la perception active. De même [Henderson 03b] s'intéresse aux problèmes spécifiques du contrôle du regard lors de la perception de scènes naturelles.

Dans [Cavanagh 04], la notion de routine visuelle est généralisée à celle de *routine attentionnelle* partant du constat que les expérimentations menées par Ullman reposaient sur des tâches difficilement accessibles à un niveau conscient. Il se base sur cette notion de conscience des traitements pour définir trois niveaux de routines visuelles :

- ▷ *Vision Routines* – Ce sont des traitements complètement inconscients comme par exemple la détection de caractéristiques particulières dans le champ visuel.
- ▷ *Attention Routines* – Ce sont des traitements déclenchés volontairement mais dont le déroulement est inaccessible à la conscience, seul le résultat est connu.
- ▷ *Cognition Routines* – Ce sont les actions que nous effectuons, la tâche qui requiert une perception visuelle. Ces routines sont constituées de plusieurs étapes conscientes au sens où elle peuvent être rapportées par le sujet qui effectue la routine.

[Ballard 95] propose une notion de routines visuelles sensiblement différente au sens où il considère le processus de perception à une échelle temporelle relativement importante (de l'ordre de quelques secondes). A une telle échelle, la vision peut être modélisée à l'aide de deux types de routines primitives : le « Quoi ? » et le « Où ? ». Les auteurs proposent de modéliser différentes tâches de perception en composant ces deux types de routines. Ces deux types de routines reposent sur une représentation des objets dans un espace d'attributs.

Des travaux se sont intéressés à la création automatique des routines visuelles en faisant appel à la programmation génétique dans le cadre de la détection de parties du corps humain (les mains plus précisément) [Johnson 94; Johnson 96]. Des opérateurs sont organisés dans un programme sous la forme d'un arbre. Ces opérateurs incluent notamment de la détection de contour et des mécanismes permettant le filtrage de listes de points. Le *fitness* d'un programme est déterminé par rapport à un jeu de données créé manuellement.

Dans [Pisan 95] les auteurs utilisent des routines visuelles afin d'interpréter des graphes. Une question symbolique y est traduite en termes géométriques et les routines sont alors utilisées afin de récupérer les informations nécessaires à la réponse. [Halelamien 04] présente quand à lui une utilisation concrète des routines visuelles pour la perception d'un robot et met en avant comment cette utilisation facilite la programmation du système perceptif. Dans [Yi 05] la notion de routine visuelle est également utilisée afin de gérer le comportement sensorimoteur d'un humanoïde virtuel. Dans [Reece 95] des routines visuelles sont mises en œuvre afin de permettre à un robot conducteur de percevoir les informations nécessaires à son évolution dans l'environnement. Toujours dans le cadre de la conduite, [Herviou 06] intègre les routines visuelles au sein de tâches de conduite afin de doter des agents conducteurs autonomes de capacités de perception permettant un comportement crédible. Dans [Salgian 98b; Salgian 98a] des routines visuelles sont mises en œuvre dans le cadre d'un système de conduite autonome, ici des comportements de conduite sont associés à une ou plusieurs routines elles même reposant sur l'utilisation d'opérateurs de bas niveau. Enfin dans [Mccallum 94] la nécessité d'une mémoire locale aux routines est mise en avant dans le cadre d'un comportement de poursuite de véhicule.

4.6.3 L'influence du contexte

La focalisation de l'attention visuelle n'est pas uniquement influencée par les stimuli visuels et les informations provenant de la vision précoce. Une autre influence dite descendante (*top-down*) reposant sur les connaissances et le contexte existe. Alors que l'influence ascendante permet au système de perception visuelle d'être attentif aux événements visuels (un mouvement par exemple) pouvant survenir de façon inattendue (on ne perçoit consciemment que ce à quoi l'on s'intéresse vraiment, ce qui est important pour la tâche que nous devons effectuer [Rullen 00; Triesch 03]). Cette influence haut niveau paraît évidente à tel point que nous n'en sommes généralement pas conscient [Brink 03]. Pourtant Yarbus [Yarbus 67] lors de ses études sur les mouvements de l'œil au cours de l'analyse d'une scène a mis en évidence des changements dans le parcours d'une image en fonction de la tâche demandée aux sujets (voir figure 4.8).

Le phénomène de *change blindness* durant lequel un sujet est insensible aux changements se produisant là où son attention n'est pas focalisée montre également cette influence de la tâche sur la perception au travers de l'attention visuelle [Hayhoe 02]. L'influence descendante de l'attention visuelle repose fortement sur des connaissances *a priori* (notre expérience). Lorsque nous percevons notre environnement, l'image formée sur la rétine est interprétée par notre cerveau (c'est le cas notamment pour les illusions d'optique telle que celle présentée dans la figure 4.9) [Cavanagh 99].

Henderson [Henderson 03b] distingue plusieurs types de connaissances pouvant influencer les mouvements du regard et la focalisation de l'attention visuelle :

- ▷ les informations sur l'instance de la scène observée (*episodic scene knowledge*) ;
- ▷ les informations sur le type de scène observée (*scene-schema knowledge*) ;
- ▷ les informations liées à la tâche en cours (*task-related knowledge*).

Le contexte qui influence l'attention visuelle et la perception visuelle d'une façon plus générale est néanmoins relativement mal défini. Dans [Bar 93] par exemple l'influence du contexte spatial (association d'objets à un type de scène particulier comme les voitures dans une rue par exemple) est mise en avant. Dans [Chun 00] c'est le contexte objet (co-occurrence d'objets, le fait de trouver une cuillère près d'une tasse de café par exemple) ainsi que le contexte dynamique (mouvement et prédiction de trajectoires, considérer par exemple le fait qu'une voiture se déplace sur la route) qui sont présentés.

Au delà des connaissances qui influencent la focalisation de l'attention visuelle, il reste le problème de la façon dont cette influence est mise en œuvre. Des travaux se sont intéressés à ce problème spécifique et ont permis la mise en évidence de cette influence sans pour autant en expliquer la nature [Connor 04]. Ces différentes études sont au delà de notre problématique et nous ne nous étendrons pas plus sur ce point. Il est malgré tout important de noter que les influences ascendantes de l'attention visuelle sont nombreuses. Elles peuvent aller jusqu'à modifier la vision lors des premiers instants de la perception même si celle-ci reste plus visible lors des traitements de haut niveau de la hiérarchie du cortex visuel [Treue 03]. Les deux influences de l'attention visuelle apparaissent en fait comme intimement liées au sein d'une

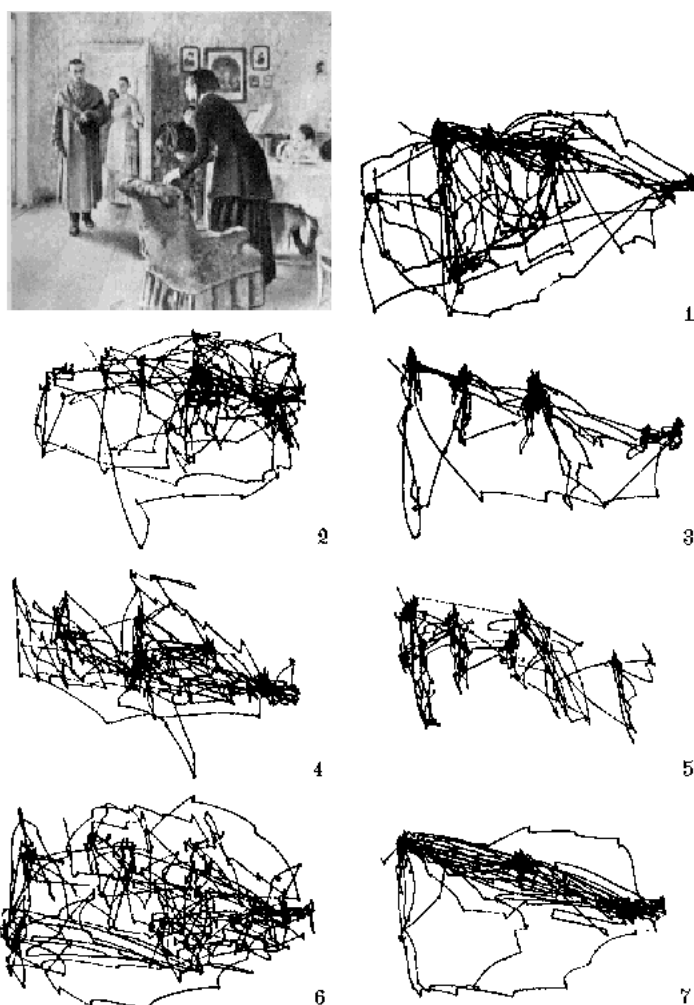


FIG. 4.8 – Parcours de l'œil dans une image et ses variations en fonction de la tâche [Yarbus 67]. Dans chaque trace, les sujets devaient examiner l'image : 1 – sans but particulier ; 2 – pour estimer le niveau de vie des personnages ; 3 – pour estimer l'âge des personnages ; 4 – pour deviner ce que faisaient les personnages avant l'arrivée des visiteurs ; 5 – mémoriser les vêtements des personnages ; 6 – mémoriser la position des objets et personnages dans la pièce ; 7 – estimer le temps depuis la dernière visite des invités.

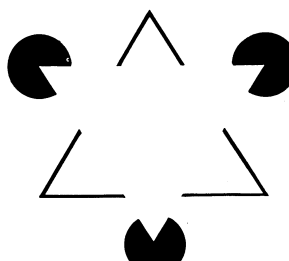


FIG. 4.9 – Image montrant le rôle de l'interprétation dans la perception visuelle.

boucle, le contexte permettant de guider la perception et les objets perçus venant compléter le contexte [Bar 04].

Pour les modèles pré-attentifs basés sur la notion de cartes de saillance, [Henderson 03b] distingue deux approches :

- ▷ Modulation d'une carte de saillance existante.
- ▷ Modification du processus de création de la carte de saillance.

La première approche est certainement la plus utilisée. Dans [Navalpakkam 02; Navalpakkam 05], une carte attentionnelle ainsi qu'une carte d'inhibition sont créées et combinées à la carte de saillance classique (modèle de Itti et Koch) par simple pondération. La carte attentionnelle repose sur des connaissances *a priori* décrivant par exemple la co-occurrence d'objets. Dans [Olivia 03] la carte de saillance est modulée par une carte décrivant les zones susceptibles de contenir un objet recherché. [Gaborski 03] présente un mécanisme de modulation de carte de saillance par une carte attentionnelle produite par un réseau de neurones. Ce réseau est entraîné afin de localiser les zones d'une image contenant les couleurs rouge, verte, bleu et jaune. Lors de son fonctionnement il permet de produire une carte attentionnelle mettant en avant les zones de l'image d'entrée contenant ces couleurs. Dans [Choi 03] une carte de saillance est inhibée en certains endroit par une carte attentionnelle permettant de déterminer les zones possédant des caractéristiques non souhaitées. Ces zones sont tout d'abord identifiées par un opérateur humain, le système étant ensuite capable de se souvenir de ces informations.

Pour [Milanese 94] c'est la seconde approche qui est utilisée par pondération des différentes cartes d'aspect dans le calcul de la carte de saillance finale. Dans [Veyret 04c] des cartes de saillance simplifiées (elles ne contiennent qu'une partie des attributs considérés dans le cadre du modèle de Koch et Ullman) sont calculées au besoin afin de vérifier des hypothèses formulées par un réseau Bayésien. Ce réseau permet de modéliser des connaissances relatives à des situations routière à identifier à l'aide d'une notion de « faits » tels que la présence d'un panneau rouge ou la présence d'un feu tricolore. Chaque fait correspond à un ou plusieurs nœuds dans le réseau, les nœuds terminaux étant associés à des vecteurs d'information pouvant être extrait de l'image d'entrée. Une boucle d'interaction est alors mise en place, les informations extraites de l'image permettant de modifier les croyances modélisées au sein du réseau bayésien et le réseau permettant de sélectionner les traitements à effectuer en fonction de la situation la plus probable à un instant donné. Enfin [Baccon 00] présente un mécanisme d'apprentissage permettant la mise en place d'une influence descendante capable de modifier le processus de création de la carte de saillance et une mécanisme de retour permettant d'influer sur l'intégration des différentes cartes d'attribut.

On trouve également des approches ne prenant en compte que le contexte comme dans [Baluja 95] où un réseau de neurones est responsable du calcul d'une carte de saillance à partir des connaissances qu'il a acquises par apprentissage. Le réseau permet de mettre en avant les zones d'image correspondant à une route, agissant à la manière d'un filtre. Dans [Veyret 04b] et [Veyret 04a] un état de l'art est présenté en ce qui concerne les approches ascendantes, descendantes et hybrides de l'attention visuelle en informatique.

4.6.4 Attention et représentation

Un aspect important de la perception visuelle humaine est le fait que l'on a à tout moment l'impression de posséder une représentation complète et précise de l'environnement. Cette impression fait supposer que l'on dispose d'une mémoire complète dans laquelle nous reconstituons totalement l'environnement qui nous entoure. Néanmoins des études sur le phénomène de *change blindness* qui se caractérise par une insensibilité aux changements lorsque l'attention n'est pas focalisée sur l'endroit où se produit ce changement plaident en faveur d'une mémoire relativement limitée lors des premiers instants de la perception visuelle [Henderson 03b; Henderson 03a; Rensink 97; Hayhoe 02].

La théorie de la cohérence attentionnelle proposée par Rensink [Rensink 00] présente une notion de mémoire partielle de l'environnement pour tenter d'expliquer plusieurs phénomènes perceptifs dont notamment le *change blindness*. Dans cette théorie, des "proto-objets" sont continuellement produits en parallèle sur la totalité du champ visuel. Ces objets, même s'ils peuvent s'avérer complexes, ne possèdent pas de cohérence spatiale ou temporelle. L'attention visuelle permet de doter une petite partie de ces « proto-objets » de la cohérence qui leur manque et ainsi de les transformer en un objet stable. Cette cohérence permet par exemple d'associer un nouveau stimuli visuel comme étant un changement d'un objet stable existant plutôt que l'apparition d'un nouveau « proto-objet ». Lorsque l'attention accordée à un objet est relâchée, celui-ci perd sa cohérence et redevient les différents « proto-objets » qui le constitue. Ainsi selon la théorie de Rensink, l'attention ne pouvant être portée que sur un nombre limité d'objets à un instant donné, la mémoire consisterait simplement en une représentation partielle de l'environnement ne contenant que les objets nécessaires à un instant donné (*just-in-time representation*).

Dans [de Fockert 01] un lien direct entre la mémoire de travail et le contrôle de l'attention visuelle est mis en évidence. Les auteurs montrent un lien entre la charge courante de la mémoire de travail et la sensibilité à des stimuli sans rapport avec une tâche spécifique (des distracteurs). Ils expliquent ce phénomène par l'attribution d'une priorité aux différentes informations présentes dans la mémoire de travail.

Dans [Brink 03] la question des objets de l'attention (d'une façon générale et pas seulement l'attention visuelle) est abordée. Il apparaît que ceux-ci sont variés et fortement dépendants de la tâche ainsi que du niveau d'observation auquel on se place. L'attention ne pouvant être attirée que par des objets concrets, qui possèdent des propriétés. Il distingue ainsi les objets attentionnels suivant différents niveaux de la perception :

- ▷ des régions de l'espace visuel ;
- ▷ des représentation *pragmatiques* qui consiste en ces zones de l'espaces auxquelles sont ajoutées des informations relatives à la tâche en cours (texture, taille, couleur, ...) ;
- ▷ des percepts tridimensionnels qui atteignent le niveau conscient.

Il est cependant important de noter que l'auteur ne voit en ces objets qu'une interprétation possible et accepte la possibilité d'interprétations différentes.

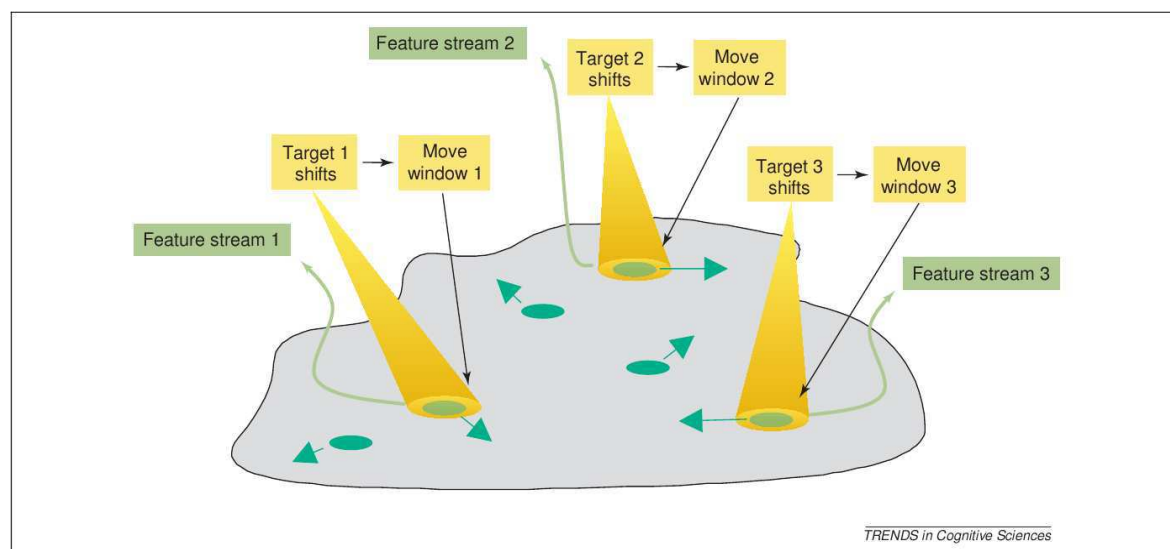


FIG. 4.10 – Le modèle attentionnel multifocal [Cavanagh 05].

Le problème de la mémoire et de son lien avec l'attention visuelle a également été étudié au travers du suivi de cibles multiples [Cavanagh 05]. Plusieurs modèles d'attention multiple ont été proposés depuis les premières expérimentations mettant en évidence ce phénomène [Pylyshyn 89]. Ici nous présentons trois de ces modèles qui supposent l'existence de multiples focus attentionnels : l'attention multifocale [Cavanagh 05], les index pré-attentifs (*FINSTs*) [Pylyshyn 89; Pylyshyn 94] et les *object files* [Kahneman 92; Wolfe 97]. Une revue plus complète des différents modèles existants est néanmoins proposée dans [Oksama 04].

Le modèle d'attention multifocale suppose plusieurs focus attentionnels opérant en parallèle. Ces différents focus restent associés à une même cible au cours de son évolution (figure 4.10). Chacun de ces focus travaille sur une fenêtre spatiale à laquelle est associé un flux capable de transférer les informations aux traitements de plus haut niveau.

Le modèle des index pré-attentifs (*FINSTs* pour *Fingers INSTantiation*) repose sur la notion d'index associés aux cibles et capables de les suivre lorsque celles-ci se déplacent. Un tel indice peut décrire une position dans l'espace aussi bien qu'une conjonction de propriétés particulières et se trouve de ce fait être relativement général. Ces index ne sont pas spécifiquement associés à l'attention visuelle mais celle-ci permet un accès plus rapide à leur propriétés. Néanmoins, ce modèle considère que l'attention ne peut être focalisée que sur un index à la fois. Le problème de l'indexation spatiale est également abordé dans [Leslie 98].

Enfin, la notion d'*object files* est proche de l'idée d'attention basée objet. Dans ce modèle les propriétés des objets suivis y sont associées sous la forme d'un « dossier » (figure 4.11). Ce « dossier » permet d'accumuler des informations sur une cible au cours de son évolution. Ses fonctions sont de sélectionner, suivre et coder des objets. Ces fonctions correspondent en fait à celles classiquement attribuées à l'attention visuelle. Ici les informations sont maintenues séparément pour les différents objets suivis. Dans [Cavanagh 05], la notion d'*object files* est considérée comme étant un mécanisme interne du modèle multifocal plutôt qu'une théorie à part entière.

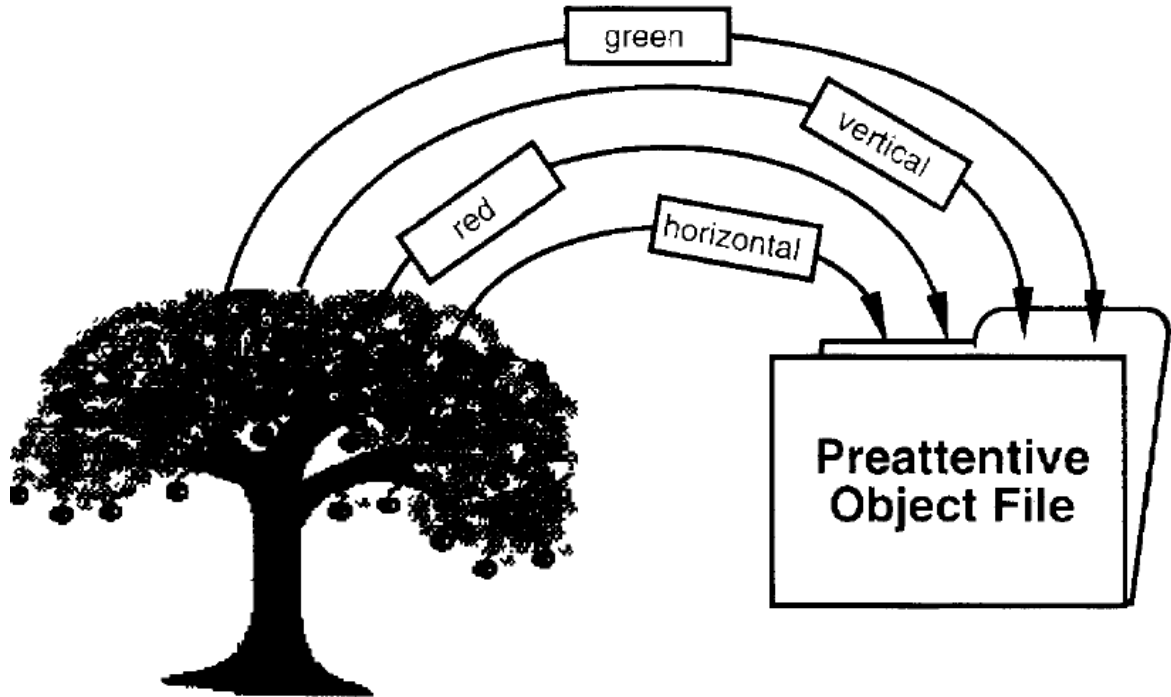


FIG. 4.11 – Représentation des *object files* [Kahneman 92; Wolfe 97].

[Backer 03] propose ce que l'on peut considérer comme une mise en œuvre de la notion d'attention multifocale. Partant des limitations évidentes des modèles classiques d'attention (inhibition spatiale plutôt que basée objet, impossibilité d'associer les informations extraites aux objets et non prise en compte d'un environnement dynamique) les auteurs proposent un nouveau modèle intégrant la notion de *FINST* au sein d'une étape semi-attentive. Ainsi des « champs de neurones » (*neural fields*) multiples, spécialisés pour différentes caractéristiques, permettent de mettre en évidence des *clusters* d'activité au sein de la carte de saillance, chacun étant associé à ce que les auteurs nomment *object files* (des *FINSTs* en réalité). Les *clusters* correspondent ainsi à un objet visuel, une partie d'un objet ou une collection d'objets. Un comportement d'exploration est ensuite en mesure de sélectionner un *cluster* particulier afin d'y effectuer des traitements de plus haut niveau tels que de la reconnaissance.

4.6.5 Contrôle attentionnel

Nous nous intéressons ici plus particulièrement aux mécanismes attentionnels⁸ dans le domaine informatique.

L'utilisation première et principale de l'attention visuelle pour le contrôle dans les

⁸ Nous parlons ici de mécanismes attentionnels au sens de l'attention visuelle humaine et non pas simplement au sens de mécanismes de focalisation tels que ceux envisagés dans la section 4.5.5.

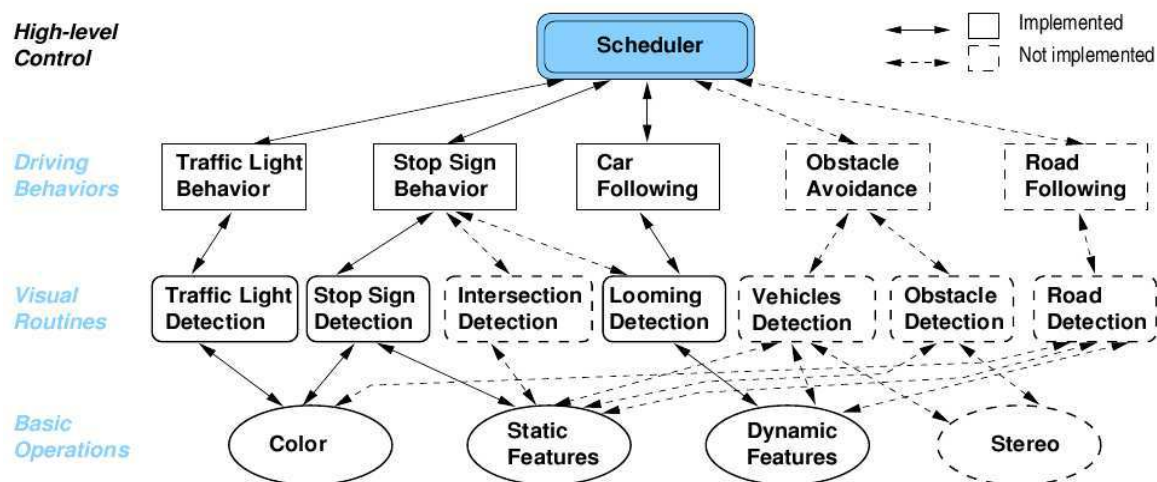


FIG. 4.12 – Organisation des routines visuelles dans [Salgian 98a].

systèmes de vision repose sur les modèles de vision précoce et la notion de carte de saillance, généralement pour offrir un point de départ au reste du système de vision (un région d'intérêt) [Paletta 05]. Des travaux tels que ceux présentés dans [Frintrop 04; Gao 04; Walther 02; Walther 04a; Walther 05; Herpers 98] sont typiques de ce genre d'utilisation. De même dans [Concepcion 96] où les mécanismes attentionnels sont utilisés afin de déterminer quelles zones de l'image doivent être traitées, ces zones subissant ensuite une classification grossière afin de déterminer si elles doivent ou non subir des traitements plus fins. Dans [Edgington 05; Edgington 04] ce type d'utilisation est mis en œuvre dans le cadre du suivi et de la classification d'animaux dans des vidéos sous-marines. L'aspect descendant de l'attention visuelle a trouvé sa place principalement dans des tâches de recherche visuelle et de reconnaissance, les différents travaux tendant à se focaliser sur l'interaction entre les deux influences de l'attention visuelle (cf. section 4.6.3).

Un autre aspect de l'attention visuelle humaine que l'on retrouve dans le cadre des systèmes de vision est celui des traitements séquentiels et la notion de routines visuelles. Plusieurs travaux dans le domaine informatique se sont intéressés à l'utilisation de telles routines et à la façon de les organiser au sein d'un système complet de perception.

Dans [Hewett 01] un système de focalisation de l'attention perceptive généraliste (fait pour tout type de perception) est présenté. L'auteur cherche à y reproduire les mécanismes attentionnels humains sans pour autant chercher à en copier l'architecture et le fonctionnement. Il considère certains mécanismes de l'attention visuelle tels que l'attention soutenue (garder un objectif au cours du temps comme dans le cadre du suivi de cibles par exemple), l'attention sélective qui permet de sélectionner une partie des données images possédant certaines propriétés et l'attention divisée qu'il considère comme essentielle pour permettre la poursuite de plusieurs tâches perceptives simultanément. Il propose ensuite un système dans le cadre d'une application au domaine de la robotique qui vise à la mise en place d'un système multimodal exécuté sur un système mono-processeur multi-tâche temps réel. Ce système est concentré autour d'un but de maintien de la cohérence d'une mémoire à court terme au travers de l'activation, l'arrêt et l'ordonnancement de routines de détection et de suivi au moyen d'un

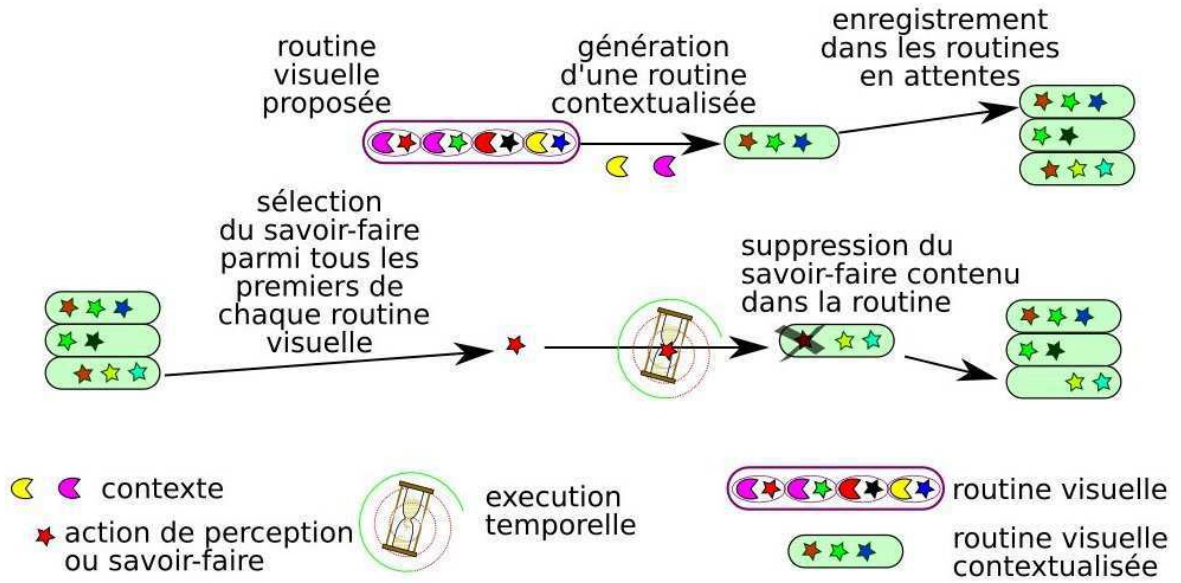


FIG. 4.13 – Contextualisation et exécution des routines visuelles dans [Herviou 06].

gestionnaire d'attention centralisé. [Sprague 01] propose une autre solution pour le même type de problème. Les auteurs présentent une architecture de contrôle basée comportement pour sélectionner et ordonnancer l'exécution des routines visuelles afin de supporter le comportement d'un humanoïde virtuel. Ici les comportements sont vus comme des processus concurrents travaillant sur des ressources partagées : les ressources de perception (routines visuelles) et les ressources d'action (contrôle de l'humanoïde). Un système d'enchères permet de sélectionner les comportements auxquels les ressources doivent être allouées à un instant donné.

Dans [Salgian 98b; Salgian 98a] les routines, utilisées dans le cadre d'un robot conducteur, sont associées à des comportements et basées sur un ensemble d'opérateurs simples (figure 4.12). Un contrôleur central se charge de sélectionner les comportements à exécuter (exécution séquentielle parmi une liste de comportements actifs), chaque comportement étant mis en œuvre par une machine à état qui se charge d'exécuter les routines visuelles nécessaires. Dans [Herviou 06] une idée similaire est développée dans le cadre de la perception de conducteurs virtuels. Ici les routines sont déclenchées au sein de tâches de conduite et leur exécution est placée sous la responsabilité d'un contrôleur de perception qui peut décider de les exécuter ou non. Ce contrôleur procède à la contextualisation des routines visuelles au travers du filtrage des actions qui la compose (figure 4.13). L'exécution de plusieurs routines se fait de façon entrelacée afin de permettre la perception de différentes informations en parallèle. Il est important de noter qu'ici les routines visuelles correspondent à des requêtes de perception sur un capteur virtuel et non pas à la mise en œuvre de traitements d'images.

4.6.6 Bilan

Dans cette section nous nous sommes intéressés à l'attention visuelle humaine ainsi qu'à ses applications et modélisations dans le domaine informatique. Nous avons décrit les caractéris-

tiques de la vision précoce et son rôle dans la sélection des stimuli visuels au travers notamment de la notion de saillance. Ensuite nous avons présenté les routines visuelles et les traitements séquentiels nécessaires à l'extraction d'informations complexes. Nous avons présenté les aspects contextuels de l'attention visuelle et montré comment les modèles à base de saillance incorporent cette influence. Enfin nous nous sommes intéressés aux liens entre l'attention visuelle et la mémoire notamment au travers des problèmes de suivi de cibles multiples.

Parmi les travaux que nous avons présentés beaucoup s'attachent à modéliser et simuler l'attention visuelle. Notre objectif est ici sensiblement différent au sens où nous cherchons à utiliser des mécanismes attentionnels afin de guider le processus de perception, de lui permettre de s'adapter aux besoins d'un processus décisionnel dont l'objectif évolue au cours du temps tout en tenant compte de ce qui peut être perçu dans l'environnement.

La notion de saillance et ses capacités de filtrage de l'information en font une composante essentielle des systèmes de perception, néanmoins il ne nous semble pas nécessaire dans le cadre de nos travaux d'effectuer le calcul d'une carte prédéfinie à chaque instant de la perception. De même les propriétés devant être incluses dans cette carte de saillance ne semblent pas si évidentes lorsque l'on ne cherche pas à simuler le système attentionnel humain.

L'influence du contexte et la notion de routine visuelle (ou stratégie de perception) apparaissent comme deux autres aspects importants de l'attention visuelle. Reste malgré tout le problème du choix et de l'ordonnancement de ces stratégies. Pour ce qui est des routines visuelles, ici encore il ne nous semble pas nécessaire dans le cadre de nos travaux de limiter le nombre d'opérateurs possibles à ceux proposés par Ullman, le nombre et la variété des algorithmes de traitement d'image existants offrant beaucoup plus de possibilités.

En ce qui concerne les liens entre l'attention visuelle et la mémoire, nous avons présenté un certain nombre de travaux et théories se basant sur l'étude du suivi de cibles multiples. Ceci nous a permis de mettre en avant différentes approches de l'attention visuelle et notamment l'attention basée objet. Les notions d'*object file* et de *FINST* nous semblent intéressantes du point de vue de la représentation des informations (propriétés associées à des objets attentionnels). Ici encore d'un point de vue informatique, ces théories ne nous paraissent pas exclusives, les *object files* mettant l'accent sur la notion d'objet alors que les *FINSTs* sont orientés sur l'idée de propriété visuelle.

Nous avons enfin présenté rapidement des exemples d'utilisation de l'attention visuelle dans le domaine informatique et plus particulièrement des utilisations de la notion de saillance et des routines visuelles.

4.7 Conclusion

Dans ce chapitre nous nous sommes intéressés au problème de la perception visuelle. Nous avons présenté les contraintes inhérentes au cadre de nos travaux, et notamment : l'utilisation de caméras vidéo et de ce fait la nécessité d'employer des techniques de vision par ordinateur, des contraintes temps réel fortes et un environnement complexe peuplé d'entités autonomes dont les conditions d'observation ne sont pas contrôlées.

Nous avons ensuite présenté différentes approches de la perception visuelle d'un point de vue théorique. Nous avons identifié principalement trois approches :

- ▷ la perception directe qui considère l'environnement comme sa propre représentation ;
- ▷ la perception indirecte qui considère que le rôle de la perception visuelle réside en une reconstruction tridimensionnelle complète de l'environnement ;
- ▷ la perception interactive dans laquelle percevoir l'environnement consiste en une adaptation des mécanismes perceptifs par rapport aux besoins et/ou attentes à l'aide de mécanismes attentionnels.

Nous avons terminé cette partie de l'état de l'art par la présentation de différents modèles de mémoire et des approches pour la représentation de l'environnement. Dans le cadre de nos travaux et étant donné les traitements d'image qu'il est nécessaire d'effectuer, il nous semble inévitable de disposer d'une représentation de l'environnement. Il reste encore à déterminer quel type de représentation et comment la construire.

Considérant le contexte de notre étude (cf. section 4.2) nous nous sommes ensuite intéressés aux différentes approches de la vision artificielle. Nous nous sommes ici orientés sur la notion de perception active consistant en une utilisation active de capteurs passifs. Parmi ces approches nous avons ensuite développé celles basées sur le contrôle des algorithmes et de leurs paramètres (perception active et vision intentionnelle) en présentant les différents mécanismes de contrôle mis en œuvre dans des systèmes de vision existants. Il ressort de cette présentation des travaux concernant la vision artificielle une légère divergence du point de vue des objectifs des systèmes de vision par rapport au cadre de notre étude, ces systèmes envisageant la vision comme un but alors que nous considérons celle-ci comme un besoin. Néanmoins les architectures modulaires utilisées dans le cadre de ces travaux nous semblent appropriées à la mise en œuvre d'un système de vision tel que nous l'envisageons.

Partant de ce constat de divergence et des capacités adaptatives du système visuel humain, nous nous sommes enfin intéressés aux mécanismes attentionnels de la perception humaine et aux modèles informatiques de ces mécanismes. Nous nous sommes concentrés sur trois aspects principaux : la vision précoce, les traitements séquentiels et l'influence du contexte. Ceci nous a permis de mettre en avant quelques notions qui nous paraissent intéressantes dans le cadre de nos travaux : les cartes de saillance qui offrent un mécanisme de filtrage des informations visuelles relativement puissant et permettent de sélectionner uniquement une petite partie des stimuli visuels et la notion de routine visuelle au sens de stratégie d'extraction d'informations. Nous avons ensuite présenté les liens entre les mécanismes attentionnels et la mémoire en mettant en avant la notion de focus attentionnel multiple. Ceci nous a permis de mettre en avant la notion d'objet d'attention au sens des objets visuels (*object files*) et des propriétés visuelles pouvant être extraites des images (*FINSTs*). Enfin nous nous sommes intéressés aux utilisations informatiques de la notion de routine visuelle et de stratégie de prise d'information et nous avons présenté différentes approches pour la sélection et l'ordonnancement de ces stratégies. De cette présentation ressort la notion de contrôleur attentionnel capable de mettre en œuvre des stratégies de perception à travers l'organisation, l'exécution et éventuellement la contextualisation de routines visuelles existantes (elles ne sont pas créées à la volée).

Chapitre 5

Le contrôle de l'attention visuelle pour construire une représentation

5.1 Introduction

Dans le chapitre 3 nous avons laissé de côté le problème de la perception de l'environnement réel par le guide virtuel, supposant que celui-ci disposait à tout instant d'une représentation accessible à la manière d'une base de données (cf. section 3.3.3).

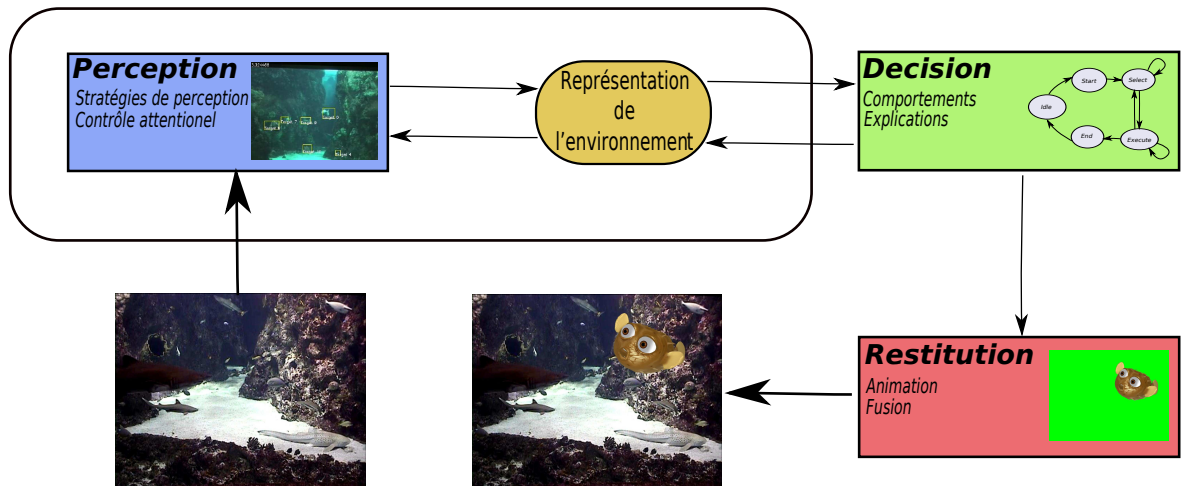


FIG. 5.1 – Positionnement de ce chapitre dans l'architecture générale de notre proposition.

Dans ce chapitre nous revenons sur ce problème de la perception visuelle (figure 5.1). Nous justifions tout d'abord le choix de l'intégration de deux type différents de perception au sein du système complet. Puis nous nous attardons sur les problèmes spécifiques au type d'environnement auquel nous nous intéressons avant de présenter l'architecture du système de perception visuelle. Cette présentation s'articule autour de deux aspects : les informations

manipulées par le système et l'exécution du système ainsi que les mécanismes de contrôle mis en œuvre. Enfin nous présentons la mise en œuvre du modèle proposé pour la construction d'une représentation d'un environnement dynamique exploitable par le processus de prise de décision du guide virtuel.

5.2 Une représentation partielle de l'environnement

Pour qu'il soit en mesure de sélectionner les explications à fournir au visiteur ainsi que pour être capable de naviguer de façon crédible dans l'environnement, le guide virtuel doit pouvoir à tout instant accéder à l'état de cet environnement ainsi qu'à celui des entités qui le peuplent.

La perception de l'environnement est effectuée par un ensemble de caméras vidéos qui fournissent un flux d'images au système de perception visuelle (cf. section 4.2). Percevoir l'état de l'environnement à partir de ces images implique l'utilisation d'algorithmes de traitement d'image coûteux en terme de temps de calcul. Les temps nécessaires à l'exécution de ces traitements empêchent leur utilisation au moment où les informations qu'il permettent de calculer sont nécessitées par le guide virtuel. Une telle utilisation impliquerait inévitablement un décalage par rapport à l'état réel de l'environnement.

Il apparaît indispensable que le guide puisse disposer d'une représentation de l'état de l'environnement afin de pouvoir prendre des décisions quant au déroulement de la visite guidée et à l'exécution de son comportement. Nous considérons que le rôle du système de perception visuelle est de construire et mettre à jour cette représentation à l'aide de mécanismes de perception indirecte (construction de connaissances à partir des images).

Les informations présentes dans cette représentation doivent permettre plusieurs choses :

- ▷ la navigation correcte du personnage incarnant le guide dans l'environnement (évitement d'obstacles par exemple) ;
- ▷ la détection d'événements importants (apparition ou disparition d'entités dans le champ visuel des caméras) ;
- ▷ la sélection des explications à présenter aux visiteurs.

Pour cela, il est nécessaire que le système de perception soit en mesure de :

- ▷ détecter les entités visibles dans l'environnement ;
- ▷ suivre ces entités au cours du temps ;
- ▷ identifier ces entités.

Ces aspects sont détaillés dans la section 5.5.

Étant donné les contraintes temporelles, l'aspect dynamique de l'environnement et la puissance de calcul limitée dont nous disposons, il n'est pas possible de créer une représentation complète et cohérente de l'environnement en temps réel.

L'ordre de grandeur de ce « temps réel » dépend des informations que l'on souhaite extraire des stimuli visuels, de l'environnement observé et de l'activité d'explication du guide virtuel. Les contraintes affectant le système de perception visuelle ne sont pas les mêmes pour tous les traitements devant être mis en œuvre. La détection de nouvelles entités faisant leur apparition dans l'environnement n'a pas besoin d'être exécutée en permanence mais doit au contraire s'effectuer de façon intelligente dans l'espace et dans le temps en fonction des besoins pour la construction de la représentation de l'environnement. Le suivi au contraire doit s'assurer de la cohérence des informations spatiales des différentes entités perçues et nécessite pour cela l'exécution de traitements plusieurs fois par seconde¹. L'identification des entités peuplant l'environnement quant à elle peut être effectuée sur une échelle de temps beaucoup plus longue de l'ordre de une à deux secondes tout en étant contrainte par la vitesse des entités que l'on cherche à décrire.

Malgré ces différences au niveau des délais d'exécution de chacun des traitements du système de perception, ils restent tous soumis à un même problème : les temps de calcul associés aux traitements nécessaires à l'extraction d'informations à partir d'images. Cette contrainte interdit, considérant une puissance de calcul limitée, la construction d'une représentation complète de l'environnement. Une telle représentation n'est d'ailleurs pas forcément nécessaire ou souhaitable. En effet, le processus de prise de décision n'a besoin que d'une partie des entités visuelles présentes dans l'environnement à un instant donné de la visite : l'entité qu'il est éventuellement en train d'expliquer. La navigation peut nécessiter une représentation spatiale plus complète : toutes les entités nécessaires à une navigation correcte mais sans pour autant connaître leur type.

Construire une représentation partielle² de l'environnement est donc suffisant et constitue la seule alternative étant donné les contraintes auxquelles nous devons faire face. Bien qu'il soit dès lors possible de créer une telle représentation, celle-ci ne doit pas être construite à *l'aveugle*. L'apparition de nouvelles entités dans le champ visuel, la mise à jour des entités connues ou encore la détection de leur disparition sont nécessaires au bon fonctionnement du processus de décision qui repose sur un contenu cohérent de la représentation par rapport à l'état de l'environnement réel. Les traitements permettant la construction et la mise à jour de cette représentation doivent être exécutés de telle sorte que le contenu de celle-ci soit le *meilleur possible* compte tenu de la tâche courante du guide virtuel. Cette notion de *meilleure représentation possible* implique une coopération entre le système de perception visuelle et le processus de prise de décision pour la construction de la représentation de l'environnement.

Pour cela, l'architecture de perception visuelle que nous proposons ne se contente pas de construire une représentation incomplète de l'environnement mais se concentre sur la construction d'une représentation partielle *adaptée* aux besoins du processus de prise de décision. Ceci est possible grâce à la mise en œuvre de mécanismes de contrôle attentionnel permettant de guider les traitements conduisant à la construction de la représentation de l'environnement (cf. section 5.4.3). Ce contrôle est effectué par le processus de prise de décision lorsqu'il interroge la représentation de l'environnement construite par le système de perception

¹ Le minimum étant aux alentours d'une dizaine de fois par seconde pour la perception de l'aquarium marin dans le cadre de notre application. Cette valeur a été déterminée dans le cadre de notre application et dépend principalement de la vitesse des objets observés.

² Nous utilisons le terme de représentation partielle aussi bien du point de vue du nombre d'entités présentes dans cette représentation que des informations concernant ces différentes entités.

(cf. section 5.4.4).

5.3 Organisation du système de perception

5.3.1 Introduction

Le système de perception visuelle que nous proposons repose sur une organisation hiérarchique des traitements afin de construire et de maintenir à jour le contenu d'une mémoire visuelle. Cette hiérarchie se base sur la distinction entre les niveaux suivants :

- ▷ le niveau image, l'extraction d'images à partir de sources vidéos (caméras) ;
- ▷ le niveau des traitements basés image, effectués à l'aide de *cartes de propriétés* et de *routines visuelles* ;
- ▷ le niveau des traitements liés aux modifications du contenu de la mémoire regroupés sous la notion de *routines de perception*.

5.3.2 Mémoire visuelle

Les traitements effectués par le module de perception ont pour objectif de créer et mettre à jour le contenu de la mémoire visuelle. Cette mémoire représente l'état de l'environnement à un instant donné sous la forme d'un ensemble d'entités visuelles décrites par des propriétés (figure 5.2). La mémoire sert de représentation de l'environnement pour le processus de prise de décision du guide virtuel.

Entités visuelles

Une entité visuelle est une zone du champ visuel sur laquelle se porte l'attention du système de perception. Un tel « focus attentionnel » est caractérisé par :

- ▷ une étendue spatiale ;
- ▷ un ensemble de propriétés (**nom: valeur**).

L'étendue spatiale est décrite par des propriétés spécifiques nécessaires à l'existence d'une entité visuelle en mémoire. Nous désignons ces propriétés sous le terme de *propriétés minimales*. Mis à part ces propriétés spéciales il n'existe aucune contrainte quant à l'existence de telle ou telle propriété en mémoire. L'ajout de propriétés se fait de façon dynamique en fonction des traitements pouvant être effectués sur une entité donnée. Plusieurs catégories de propriétés peuvent être envisagées parmi lesquelles :

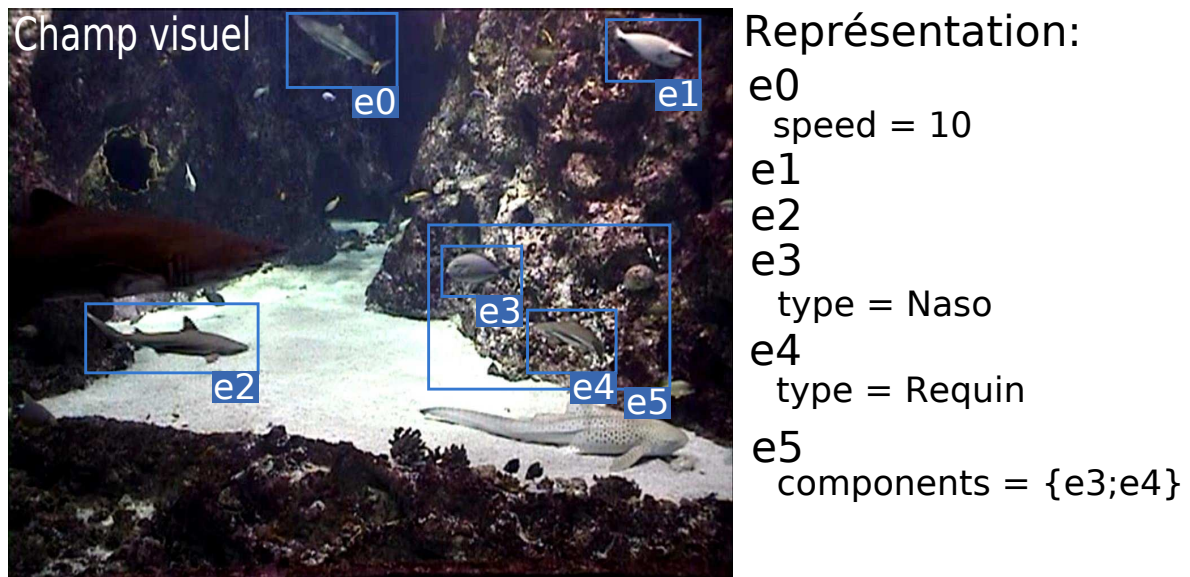


FIG. 5.2 – Mémoire visuelle. L'environnement y est représenté comme un ensemble d'entités visuelles décrites par des propriétés.

- ▷ des propriétés spatiales (position, distance) ;
- ▷ des propriétés géométriques (taille, forme) ;
- ▷ des propriétés cinématiques (direction, vitesse, trajectoire) ;
- ▷ des propriétés visuelles (couleur, texture) ;
- ▷ des propriétés sémantiques (type, rôle) ;
- ▷ des propriétés de composition (autres entités visuelles).

La figure 5.2 présente un exemple d'instance de cette mémoire visuelle. Ici six entités visuelles sont présentes ainsi que des exemples de propriétés (mis à part les propriétés minimales). Les entités $e0$, $e1$, $e2$, $e3$ et $e4$ correspondent à des objets uniques alors que l'entité $e5$ est composée de deux autres entités ($e3$ et $e4$). Il est à noter que l'entité composite $e5$ ne se limite pas à ses constituants et peut posséder des propriétés qui lui sont propres (vitesse moyenne d'un groupe d'entité, interactions en cours entre les constituants, *etc*).

Mise à jour

Un mécanisme d'oubli permet d'assurer la cohérence du contenu de la mémoire visuelle en supprimant les entités dont les propriétés minimales ne sont pas mises à jour depuis une certaine durée. A tout instant, une entité visuelle en mémoire peut se trouver dans l'un des états suivants (figure 5.3) :

- ▷ provisoire ;

- ▷ valide ;
- ▷ active ;
- ▷ obsolète.

Lors de sa création l'entité est considérée comme étant dans l'état *provisoire*. Lorsqu'une entité entre dans l'état *obsolète*, elle est retirée de la mémoire.

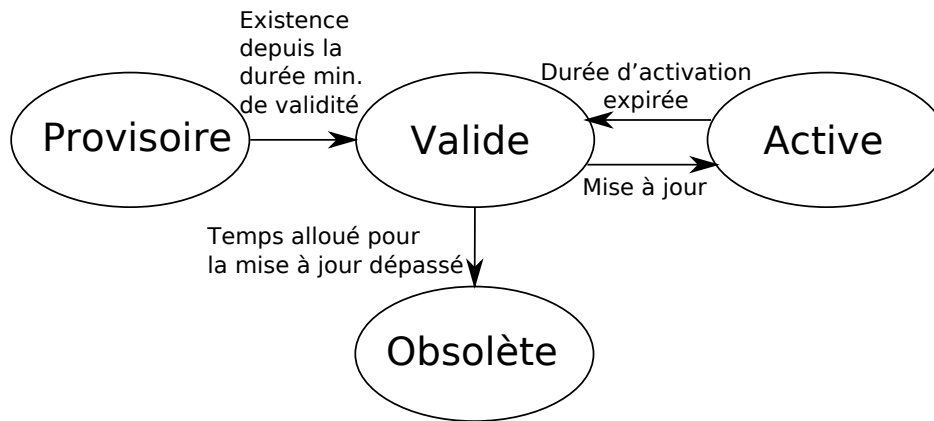


FIG. 5.3 – Les différents états possibles pour une entité visuelle.

Le passage d'une entité d'un état à un autre se fait en fonction des paramètres temporels suivants associés à chaque entité :

- ▷ durée de vie maximum ;
- ▷ durée d'activation ;
- ▷ durée minimum de validité.

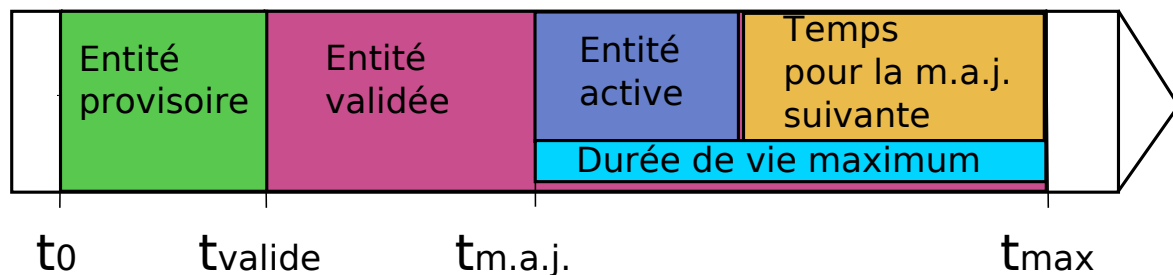


FIG. 5.4 – Paramètres de gestion de la durée de vie d'une entité visuelle.

La durée de vie maximum définit le temps au bout duquel une entité qui n'est pas mise à jour doit être retirée de la mémoire (bleu clair sur la figure 5.4). Chaque fois que les propriétés minimales d'une entité sont mises à jours, la date de suppression de cette entité est réévaluée en fonction de cette durée de vie maximum.

La durée d'activation définit la durée pendant laquelle une entité qui vient d'être mise à jour est considérée comme active, c'est à dire que les informations qui lui sont associées sont cohérentes avec l'état courant de l'environnement (bleu foncé sur la figure 5.4).

La durée minimum de validité impose une durée minimale d'existence en mémoire pour que l'entité visuelle soit considérée comme valide (zone verte sur la figure 5.4). Cette notion de validité permet de limiter l'influence de fausses détections en ne considérant que les entités présentes en mémoire depuis une durée suffisamment conséquente.

Si l'on prend par exemple les paramètres temporels suivants :

$maximumLifetime = 1.5s$

$activeTime = 0.8s$

$validityTime = 0.3s$

On peut alors envisager l'évolution suivante pour une entité visuelle donnée en mémoire :

t_0	création en mémoire avec les propriétés minimales, l'entité est provisoire
$t_1 = t_0 + 0.3$	l'entité passe de l'état <i>provisoire</i> à l'état <i>valide</i>
$t_1 < t_2 < t_1 + 1.5$	l'entité est mise à jour, elle passe dans l'état <i>active</i>
$t_3 = t_2 + 0.8$	l'entité n'est pas mise à jour depuis <i>activeTime</i> secondes, elle repasse dans l'état <i>valide</i>
$t_3 < t_4 < t_3 + 1.5$	nouvelle mise à jour de l'entité, elle repasse dans l'état <i>active</i>
$t_5 = t_4 + 0.8$	l'entité n'est pas mise à jour depuis <i>activeTime</i> secondes, elle repasse dans l'état <i>valide</i>
$t_6 = t_4 + 1.5$	l'entité n'est pas mise à jour depuis <i>maximumLifetime</i> , elle est retirée de la mémoire.

Il est important de noter qu'une entité donnée peut être sujette à plusieurs mises à jour de ses propriétés minimales avant d'être considérée comme valide. La figure 5.5 présente les courbes d'évolution du nombre d'entités dans chacun des états possible au cours du temps. Ces données ont été obtenues sur une vidéo contenant trente entités visibles en permanence. Ces entités sont en mouvement au sein d'un espace torique (une entité disparaissant sur un bord réapparaît automatiquement sur le bord opposé).

Le rôle du système de perception visuelle dans son ensemble est de construire et mettre à jour la mémoire visuelle au travers des actions suivantes :

- ▷ ajout d'une entité visuelle ;
- ▷ suppression d'une entité visuelle ;
- ▷ mise à jour de propriétés (et création de cette propriété si elle n'existe pas encore).

Différentes propriétés peuvent nécessiter des délais de mise à jour différentes. Ainsi, pour une entité visuelle en mouvement, une propriété **type** d'une entité donnée est valable durant toute sa durée de vie alors que les propriétés minimales ou des propriétés **direction** ou **vitesse** nécessiteront des mises à jour fréquentes.

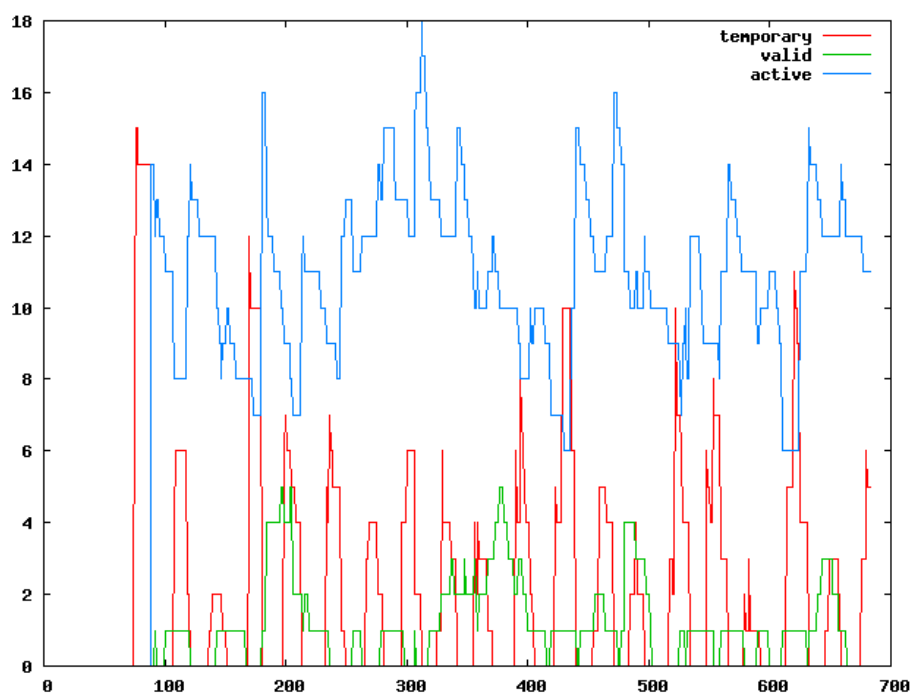


FIG. 5.5 – Évolution du nombre d'entités dans chacun des états possible : provisoire (rouge) ; valide (vert) ; active (bleu).

5.3.3 Extraction d'informations à partir des images

Le rôle du module de perception visuelle est de construire une représentation de l'environnement perçu par un ensemble de caméras vidéos. Cette représentation doit être exploitable par le module pédagogique afin qu'il soit en mesure de prendre des décisions appropriées. Pour cela, la représentation de l'environnement doit contenir des informations pertinentes (les propriétés) sur les objets visibles dans l'environnement. Cela implique non seulement d'être capable de détecter ces objets, mais également de pouvoir extraire ces informations particulières et être capable de les maintenir à jour au cours du temps.

Pour effectuer une telle tâche, nous proposons d'utiliser un module de perception automatique. Les routines de perception ont la charge de modifier le contenu de la mémoire visuelle au travers du calcul de propriétés spécifiques. Le calcul de ces propriétés peut reposer sur d'autres propriétés ou sur des traitements de bas niveau appliqués directement aux images.

L'extraction d'informations à partir des images provenant des caméras repose sur deux mécanismes complémentaires :

- ▷ les cartes de propriétés ;
- ▷ les routines visuelles.

Cartes de propriétés

La notion de carte de propriétés est inspirée de celle des cartes de saillance utilisées généralement pour la modélisation des processus de vision précoce dans le cadre de la simulation de l'attention visuelle humaine (cf. section 4.6.1).

L'idée d'une telle carte est de mettre en valeur des zones d'une image (ou séquence d'images) correspondant à une propriété visuelle (ou combinaison de propriétés) particulière. Une carte de propriétés est décrite par une fonction $f : \text{img} \rightarrow \text{map}$ qui pour une image donnée renvoie une image de mêmes dimensions. Les différents points de cette image traitée représentent une propriété ou combinaison de propriétés particulière (point en mouvement ou possédant une couleur spécifique par exemple). Le rôle d'une carte de propriétés est d'abstraire les algorithmes de traitement d'image ainsi que leur paramètres au cœur de la fonction f . Cette fonction peut éventuellement prendre des paramètres de « haut niveau » permettant d'adapter les traitements à effectuer. Les traitements effectués par la fonction f consistent en l'application successive de différents filtres élémentaires tels que :

- ▷ **motion** – un filtre mettant en évidence les zones de l'image correspondant au mouvement ;
- ▷ **red** – un filtre mettant en évidence les zones de l'image correspondant à la couleur rouge ;
- ▷ **green** – un filtre mettant en évidence les zones de l'image correspondant à la couleur verte.
- ▷ **blue** – un filtre mettant en évidence les zones de l'image correspondant à la couleur bleue.
- ▷ **yellow** – un filtre mettant en évidence les zones de l'image correspondant à la couleur jaune.
- ▷ **gabor** – un filtre mettant en évidence les zones de l'image correspondant à des arrêtes orientées.
- ▷ **depth** – un filtre mettant en évidence les zones de l'image correspondant à un intervalle de profondeur donné.

A ces opérateurs s'ajoute des traitements plus classiques tels que des opérations morphologique (érosion, dilatation, *etc*) ou de mise à l'échelle (pyramide d'images).

Il est possible de combiner les propriétés entre elles à l'aide de deux opérateurs **and** et **or** permettant respectivement de mettre en valeur des zones de l'image possédant plusieurs propriétés à la fois ou plusieurs zones possédant des propriétés différentes. Ces opérateurs sont mis en œuvre à l'aide d'opérations de multiplication et d'addition d'images respectivement. La multiplication de deux images correspond à une multiplication de chacun des points qui composent la première image par le point correspondant dans la seconde image suivi d'une étape de normalisation afin de s'assurer que les valeurs obtenues se situent dans un intervalle de valeurs valide. L'addition d'images correspond à la sélection du point d'intensité maximum

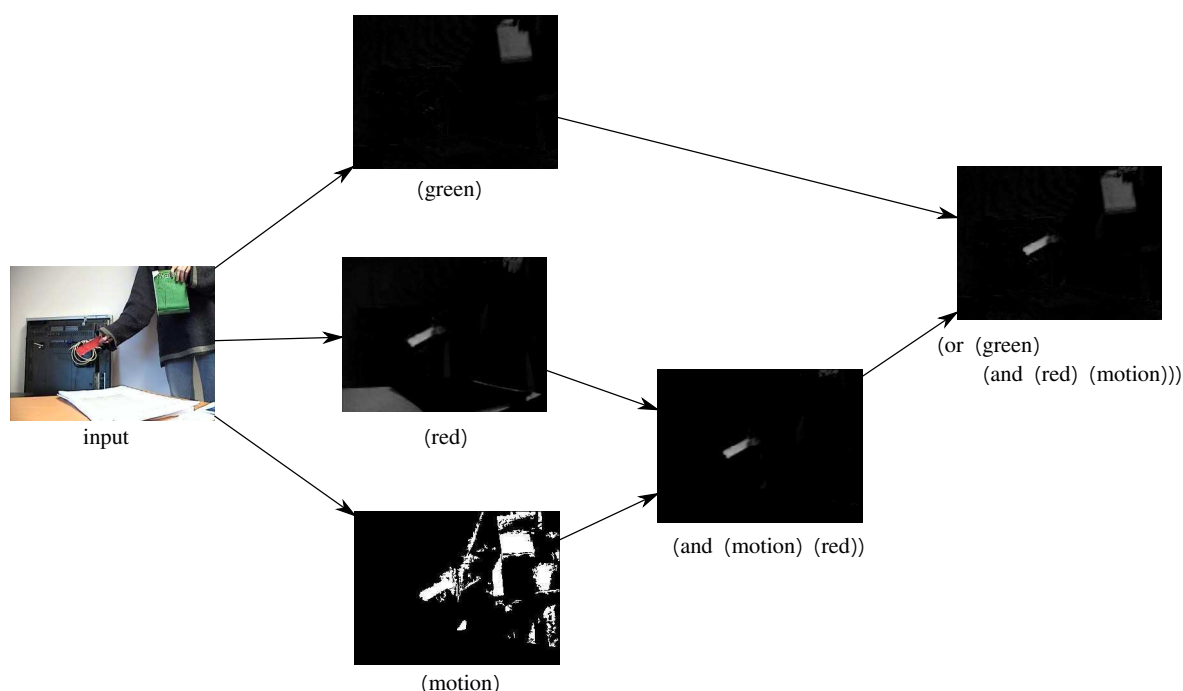


FIG. 5.6 – Exemple d'utilisation de cartes de propriétés.

parmi l'ensemble des points des images impliquées dans l'opération et ce pour chacun des points de chaque image. La définition d'une carte de propriétés données se fait à l'aide d'une expression selon la syntaxe pseudo-BNF suivante :

```

<map-exp>      ::= <or-exp> | <and-exp> | <filter-exp> ;
<or-exp>       ::= "(or" , <map-exp>+ , ")" ;
<and-exp>      ::= "(and" , <map-exp>+ , ")" ;
<filter-exp>   ::= "(" , <filter-name> , <filter-param>* , ")" ;

```

La figure 5.6 présente un exemple d'utilisation de cartes de propriétés. Les traitements sont ici effectués sur une séquence d'images contenant un objet de couleur verte et un objet de couleur rouge tenus par le manipulateur. Ce manipulateur fait bouger ces objets de telle sorte qu'un mouvement puisse être détecté dans la séquence d'images. La carte de propriétés demandée correspond à l'expression :

(or (green) (and (red) (motion)))

Une telle expression permet de mettre en valeur les zones de l'image de couleur verte ainsi que les zones rouges en mouvement. Le graphe présente la décomposition des traitements effectués lors du calcul de cette carte. Tout d'abord trois cartes simples sont calculées :

▷ la couleur verte : (green) ;

- ▷ la couleur rouge : (**red**) ;
- ▷ le mouvement : (**motion**).

Les cartes de propriétés nous permettent de mettre en évidence des zones de l'image présentant une propriété ou combinaison de propriétés particulières au travers d'expressions simples. Ce mécanisme inspiré des cartes de saillance nous offre un moyen de filtrer les stimuli visuels afin de localiser l'extraction d'informations dans l'espace image. Ici les cartes de propriétés sont calculées uniquement sur demande, les traitements qu'elles impliquent pouvant s'avérer coûteux.

Routines visuelles

L'extraction d'informations à partir d'images (traitées ou non) est effectuée à l'aide de *routines visuelles*. Les informations extraites par une routines peuvent par exemple consister en une extraction de *blobs*, le calcul d'une boîte englobante ou d'un descripteur de forme. Une routine visuelle est d'une façon générale une fonction : $g : img \rightarrow (x_0, \dots, x_n)$ qui à une image d'entrée associe une liste de valeurs. Ici les valeurs peuvent être de plusieurs types différents (entiers, réels, chaînes de caractères, ...), l'interprétation du contenu de cette liste étant à la charge de celui qui a fait la demande d'extraction d'informations et qui est censé en connaître le format. Une routine visuelle peut nécessiter des paramètres externes afin d'être spécialisée en fonction du contexte.

La figure 5.7 présente sur l'image de gauche sous forme graphique le résultat d'une routine visuelle permettant d'obtenir des informations sur la boîte englobante d'un objet dans l'image à un instant donné à partir d'un point à l'intérieur de cet objet. Cette routine nécessite en entrée les paramètres spatiaux d'une entité visuelle. À partir de ce point d'origine un calcul de composantes connexes est effectué. Le résultat de ce calcul est ensuite filtré pour éviter des informations aberrantes telles que l'extraction d'une zone comprenant l'image complète. L'image de droite de la figure 5.7 présente le résultat de l'exécution d'une routine permettant de détecter plusieurs entités à la fois. Cette routine repose sur les mêmes mécanismes que ceux décrits pour l'exemple précédent alliés à un mécanisme d'inhibition permettant de retirer les zones déjà traitées d'une image.

5.3.4 Routines de perception

Modifier le contenu de la mémoire visuelle au travers des différentes actions possibles (ajout/suppression d'entités visuelles et création/modification de propriétés) est sous la responsabilité d'un ensemble de *routines de perception*. Une routine de perception est capable de calculer une ou plusieurs propriétés pour une entité visuelle particulière. Pour cela elle peut nécessiter d'autres propriétés. Celles-ci peuvent provenir de deux types de sources :

- ▷ la mémoire (propriétés déjà calculées) ;
- ▷ de routines visuelles agissant sur des cartes de propriétés (propriétés non calculées).

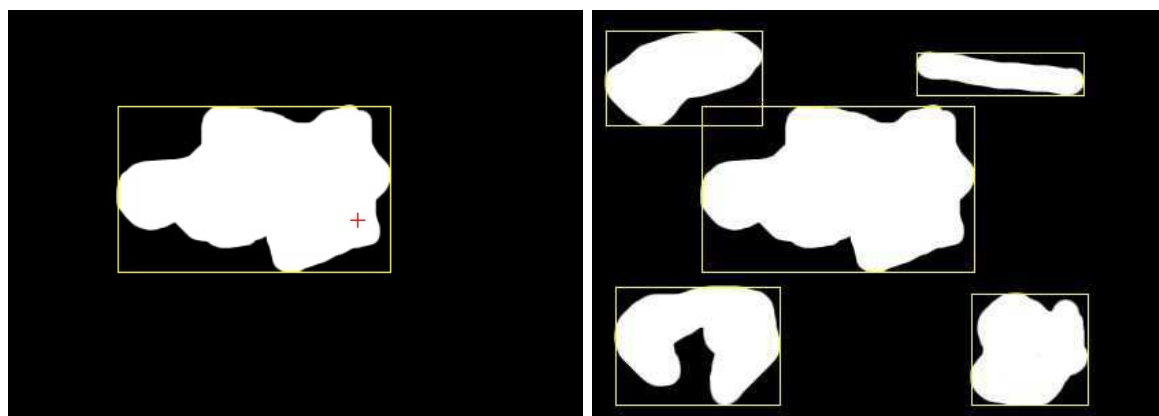


FIG. 5.7 – Résultat de l'exécution de routines visuelles.

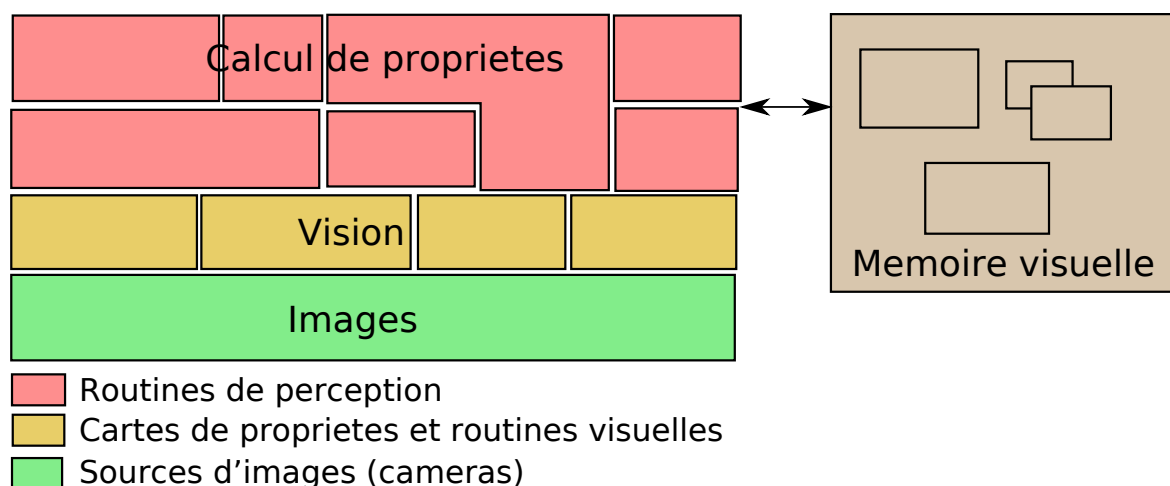


FIG. 5.8 – Organisation des traitements dans le système de perception visuelle.

Lorsqu'une routine de perception nécessite une propriété qui n'est pas présente dans la mémoire, elle a la possibilité de demander le calcul de cette propriété à la routine qui en a la charge. Les routines sont donc interdépendantes donnant naissance à un réseau dans l'organisation des traitements.

L'exécution d'une routine de perception n'assure pas l'obtention d'un résultat valide. Il est en effet possible qu'une propriété particulière que l'on cherche à calculer ne soit pas présente dans la zone du champ visuel que l'on considère. C'est par exemple le cas lorsque l'on cherche à déterminer le type d'une entité visuelle : une routine peut être spécialisée dans l'identification d'un type particulier qui s'avère ne pas être le type de l'entité considérée.

L'exécution d'une routine particulière nécessite de lui fournir un ensemble de paramètres qui sont dépendant de la routine en question. Ces paramètres peuvent définir par exemple la zone du champ visuel sur laquelle la routine doit effectuer ses traitements ou encore des contraintes sur le résultat qu'elle renvoie.

5.3.5 Bilan

Nous avons présenté l'architecture générale du système de perception visuelle que nous proposons. L'objectif de ce système est la construction ainsi que la mise à jour d'une mémoire visuelle servant de représentation de l'environnement pour le processus de prise de décision du guide virtuel.

Nous avons proposé une notion de mémoire visuelle centrée sur la notion de focus attentionnel au travers de la notion d'entité visuelle et de propriété. Les modifications du contenu de cette mémoire reposent sur un ensemble de routines de perception mettant en œuvre le calcul des propriétés des entités visuelles. Ces routines de perception peuvent nécessiter des propriétés d'entités au travers d'autres routines de perception ou l'extraction d'informations image à l'aide de routines visuelles.

L'extraction d'informations à partir des images se fait à l'aide de cartes de propriétés (filtres) et de routines visuelles (opérateurs d'extraction). Ces deux notions sont la base de tous les traitements devant être effectués sur des images dans notre système de perception. Elle permettent d'abstraire les algorithmes utilisés au profit de notions plus « génériques » et plus facilement manipulables par le reste du système décrit dans la suite de ce chapitre.

Ces différentes notions permettent d'organiser le système de perception visuelle de manière hiérarchique du point de vue des traitements à effectuer tout en proposant une représentation de l'environnement souple.

5.4 Construction d'une représentation : exécution et contrôle

5.4.1 Introduction

Après nous être intéressés à l'aspect « données » du système de perception au travers de son organisation du point de vue de la manipulation des informations et de la représentation de l'environnement reste le problème de l'exécution de ce système et des mécanismes de contrôle permettant cette exécution.

Ainsi que nous l'avons précisé dans la section 5.2, il n'est pas envisageable de traiter la totalité du champ visuel en permanence. La totalité des *routines de perception* permettant de calculer les propriétés des différentes entités visuelles présentes en mémoire ne peuvent pas être exécutées en permanence. Nous proposons ici de contrôler cette exécution avec la notion de stratégie de perception. Une telle stratégie consiste en l'organisation des traitements permettant le calcul d'une propriété donnée dans le temps et/ou l'espace en fonction d'un contexte (traitements et temps disponibles, nombre d'entités visuelles en mémoire, *etc*). Ces stratégies sont mises en œuvre par les routines de perception elles-mêmes.

Afin que la représentation construite par le système de perception soit la *meilleure possible* pour le processus de prise de décision, il est important de prendre en compte l'influence de

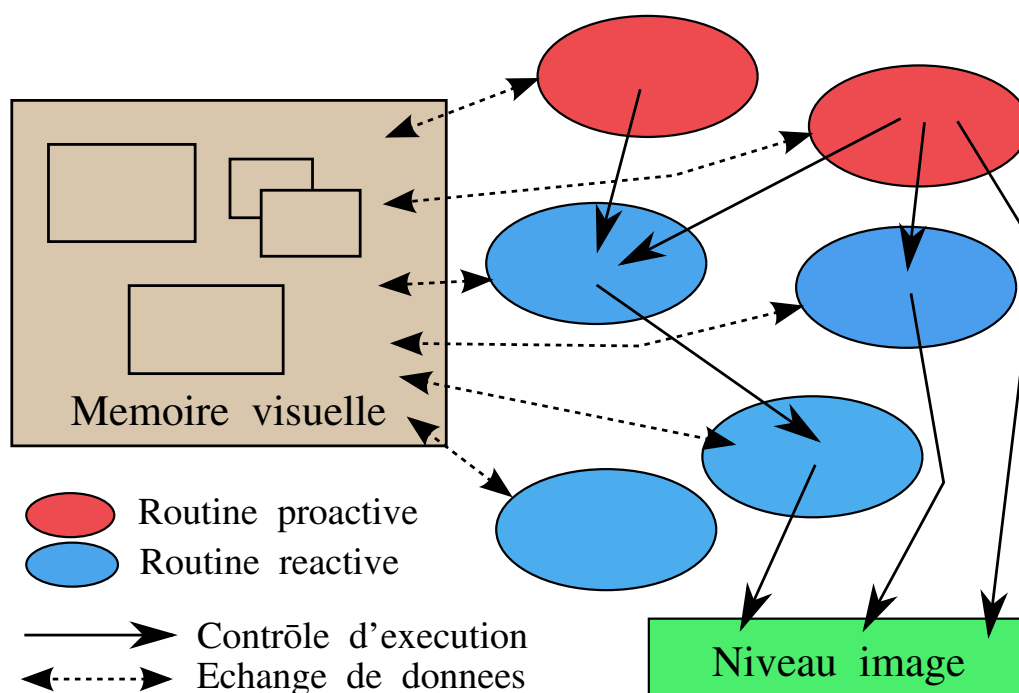


FIG. 5.9 – Exécution et contrôle au sein du système de perception visuelle.

celui-ci au sein des stratégies de perception. Cette prise en compte du processus de prise de décision s'effectue à l'aide de la notion d'intérêt qui permet de modéliser les besoins en terme de contenu de la représentation de l'environnement au cours du temps.

Le contrôle de l'intérêt au sein des routines mettant en œuvre des stratégies de perception s'effectue de manière indirecte et transparente au travers des requêtes envoyées par le processus de prise de décision à la représentation de l'environnement.

5.4.2 Stratégies de perception

La création et la maintenance du contenu de la mémoire visuelle résulte de l'exécution des différentes routines de perception existantes dans le système. Cette exécution peut se faire de deux façons :

- ▷ réactive, c'est à dire en réponse à une demande explicite de la part d'une autre routine ;
- ▷ proactive, c'est à dire selon les besoins de la routine elle même, lorsqu'elle le décide.

La figure 5.9 présente une vue schématisée de ce mode d'exécution et de contrôle. Les routines de perception proactives (en rouge) mettent en œuvre des stratégies de perception au travers du choix des routines et traitements de niveau inférieur qu'elles souhaitent déclencher. Ce choix s'effectue en fonction de l'état de la mémoire visuelle, d'un état interne et de l'intérêt

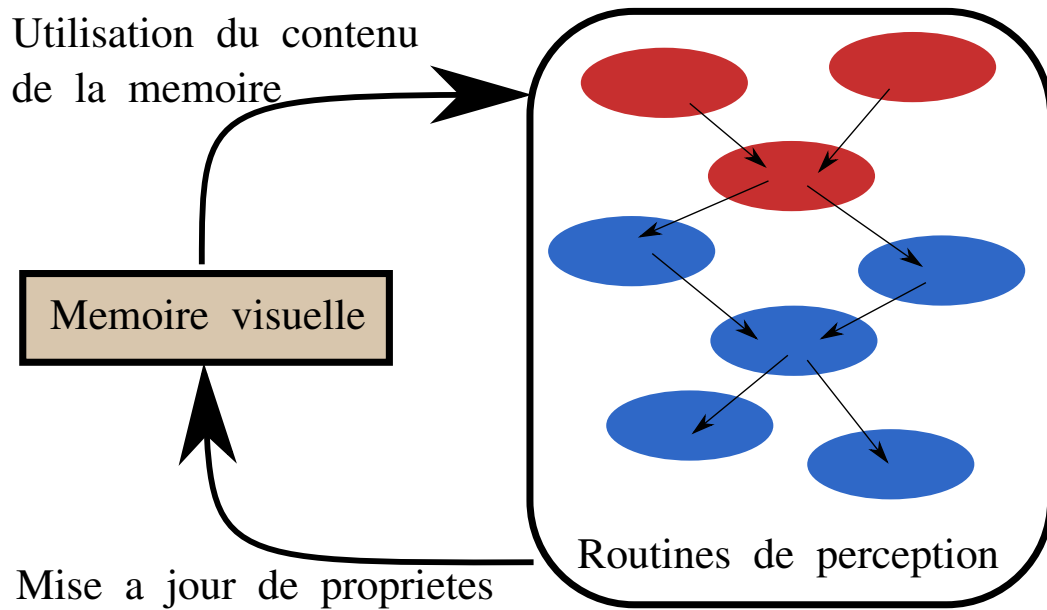


FIG. 5.10 – Boucle entre traitements et contenu de la mémoire lors de l'exécution des routines de perception.

permettant de refléter l'influence du processus de prise de décision sur le système de perception visuelle (cf. section 5.4.4).

Si une routine nécessite une propriété qui n'est pas présente en mémoire, elle demande explicitement le calcul de cette propriété à la routine qui en a la charge. Ainsi l'exécution d'une stratégie de perception donnée entraîne en cascade le déclenchement d'autres routines et traitements de niveau inférieur.

Ce mécanisme d'exécution instaure une boucle entre les deux aspects complémentaires de la perception visuelle :

- ▷ l'aspect descendant qui réside dans le déclenchement explicite de routines dans le cadre des stratégies de perception (choix des traitements en fonction de connaissances, de la mémoire visuelle et du processus de prise de décision) ;
- ▷ l'aspect ascendant que constituent les modifications de la mémoire visuelle effectuées par les routines exécutées (résultat des traitements, susceptible d'influencer la prise de décision).

Ces deux aspects entrent en interaction au sein des routines de contrôle qui mettent en œuvre les stratégies de perception en fonction de leur état interne et du contenu de la mémoire visuelle (figure 5.10).

Les différentes routines exécutées vont entraîner la mise à jour de certaines propriétés de la mémoire visuelle. Il est important de noter ici que c'est chaque routine qui a la charge de la cohérence des propriétés qu'elle est capable de calculer. Le maintien effectif de cette cohérence

dépend du type de propriété considérée, certaines pouvant nécessiter une mise à jour lors du déplacement de l'entité visuelle considérée (position par exemple), d'autre étant valable tout au long de la vie d'une entité (type par exemple).

Dans la suite de ce chapitre nous désignerons les routines proactives mettant en œuvre des stratégies de perception sous le terme de *comportement* de perception.

Le choix des routines pouvant être considérées comme étant proactives dépend au final de l'application envisagée et des routines existantes. Plusieurs critères sont néanmoins nécessaires afin que ce mécanisme présente un intérêt

- ▷ pour être en mesure de mettre en œuvre des stratégies de perception intéressantes, les routines candidates doivent disposer de choix multiples dans les traitements nécessaires au calcul de propriétés qu'elles mettent en œuvre. Une stratégie de détection par exemple doit décider des traitements à effectuer et des zones de l'image où appliquer ces traitements ;
- ▷ du point de vue du contrôle externe (cf. section 5.4.4), les stratégies de perception mises en œuvre doivent présenter un intérêt particulier, avoir un sens du point de vue de la représentation et son utilisation. Une stratégie de reconnaissance par exemple permet d'adapter les traitements en fonction du type d'entités recherchées.

5.4.3 Influence du processus de prise de décision : notion d'intérêt

L'influence de la tâche du guide virtuel sur le système de perception visuelle est modélisée au travers de la notion d'intérêt. Celle-ci permet de matérialiser la pertinence d'un aspect particulier (une propriété visuelle spécifique ou un type d'objet donné) du processus de représentation de l'environnement du point de vue du processus de prise de décision. L'intérêt entre en jeu lors du choix des routines de perception au sein des stratégies de perception offrant ainsi un mécanisme de contrôle indirect sur la construction de la représentation de l'environnement.

L'idée de la notion d'intérêt est de concrétiser la capacité attentionnelle de chaque comportement de perception. Celle-ci porte sur ce que nous nommons ici des *objets d'intérêt* qui vont permettre la mise en œuvre des stratégies de perception. Ces objets d'intérêt dépendent du comportement en question et des stratégies de perception qu'il cherche à mettre en œuvre. Un comportement de reconnaissance va par exemple proposer des objets d'intérêt sur les types, un comportement de suivi proposera des objets d'intérêt sur les entités elles-mêmes. Chaque comportement dispose d'une quantité d'intérêt fixée répartie entre tous les objets d'intérêt d'un même type. Un comportement de perception a la possibilité de mettre en avant plusieurs types d'objets d'intérêt suivant les stratégies qu'il souhaite mettre en œuvre³.

Pour chaque type d'objet d'intérêt, la totalité de l'attention disponible est répartie sur les différentes instances des objets de ce type afin de modéliser les besoins du comportement

³ C'est le cas notamment du comportement de reconnaissance présenté dans la section 5.5.3.

Algorithme 5.4.1: INTERESTATTENUATION(*interestValues*, *attenuationSpeed*, *elapsedTime*, *interestMean*)

- *interestValues*: ensemble des valeurs d'intérêt d'un type d'objet particulier
- *attenuationSpeed*: vitesse d'atténuation souhaitée
- *elapsedTime*: le temps écoulé depuis la dernière procédure d'atténuation
- *interestMean*: moyenne des valeurs d'intérêt

/ Calcul du pas d'atténuation */*

$step \leftarrow attenuationSpeed * elapsedTime$

for each $i \in interestValues$

do $\left\{ \begin{array}{l} \textit{/* Si la distance à la moyenne est supérieure au pas d'atténuation */} \\ \textit{/* on se rapproche de la moyenne de step */} \\ \textit{/* Sinon on fixe la valeur à la moyenne */} \\ \textbf{if } ABS(interestMean - i) > step \\ \quad \textbf{then } i \leftarrow i + step * \frac{interestMean - i}{ABS(interestMean - i)} \\ \quad \textbf{else } i \leftarrow interestMean \end{array} \right.$

de perception et de guider les stratégies qu'il met en œuvre. Ainsi lors du choix des routines de perception à déclencher, le comportement de perception va se baser sur deux aspects :

- ▷ le contenu courant de la mémoire visuelle ;
- ▷ la répartition de l'intérêt sur les différents objets définis.

L'influence du processus de prise de décision sur les stratégies de perception s'effectue par augmentation de l'intérêt porté sur une instance d'objet particulier. En l'absence d'influence externe, l'intérêt est réparti équitablement entre les différents objet d'intérêt existants. Lorsque l'intérêt pour un objet particulier est augmenté, l'intérêt disponible pour les autres objets diminue de façon proportionnelle. Un mécanisme d'atténuation permet de faire tendre le système vers une répartition équitable au cours du temps. Ainsi lors de chaque pas d'exécution des comportements de perception, les valeurs d'intérêt sont équilibrées selon l'algorithme 5.4.1.

Cet algorithme tend à répartir équitablement les valeurs d'intérêt sur l'ensemble des objets suivant une vitesse d'atténuation définie *a priori*. Il est à noter que cette atténuation se fait de manière discrète par rapport à un pas défini à l'avance. Lorsque l'écart entre la valeur d'intérêt à atténuer et la moyenne des intérêts est inférieur à la valeur de ce pas, alors la valeur d'intérêt est affectée à la valeur moyenne des intérêts. Cette valeur moyenne est calculée avant la procédure d'atténuation, toutes les valeurs d'intérêt étant ainsi modifiées par rapport à une même moyenne.

Un tel mécanisme permet au système de perception d'*oublier* les aspects auxquels le

processus de prise de décision s’intéresse lorsque celui-ci ne le spécifie pas explicitement. Les valeurs d’intérêt pouvant être soumises à une influence externe au système de perception, nous avons également mis en place un mécanisme de normalisation qui assure que la quantité d’attention (la quantité totale d’intérêt) au sein d’un comportement reste constante quel que soit le nombre d’aspects qu’il gère ou la « puissance » de l’influence du module pédagogique. Cette normalisation permet de limiter l’inertie des variations d’intérêt.

La vitesse d’atténuation de l’intérêt qui contrôle le rééquilibrage du système est définie en terme d’une quantité d’intérêt devant être retirée en une seconde. Ceci, allié au fait que la quantité totale d’intérêt est limitée, implique que la vitesse réelle d’atténuation est sensible au nombre d’objets d’intérêt présent à un instant donné. Plus ce nombre est grand, moins l’intérêt maximum possible sera élevé et plus l’atténuation sera rapide. La section 6.4.3 présente des expérimentations concernant ces mécanismes d’équilibre et leur paramétrage.

5.4.4 Contrôle de l’intérêt : la représentation comme interface

Ainsi que nous l’avons précisé dans la section 3.3.3, du point de vue du processus de prise de décision, la perception s’effectue de façon directe par interrogation de la représentation de l’environnement à la manière d’une base de données. Nous proposons d’influencer l’intérêt des différents comportements de perception directement au travers de ces requêtes partant de l’hypothèse que celles-ci représentent les besoins du processus de prise de décision en terme de représentation de l’environnement.

```
;; Get near sharks or all approaching entities  
(get-visual-entities (or (and (type ‘‘Shark’’)   
                             (distance < 100))   
                         (approaching)))  
;; Get all entities contained in a radius of  
;; 500 units around the guide’s position  
(get-visual-entities (around 500))
```

LISTING 5.1 – Exemples d’interrogation de la représentation de l’environnement.

Ces requêtes permettent au processus de prise de décision d’interroger la représentation de l’environnement en définissant un ensemble de contraintes qui décrivent les entités qu’il souhaite percevoir. Le listing 5.1 présente deux exemples de telles requêtes. La première permet de récupérer les entités de type “Shark” à moins de 100 unités de la position du guide virtuel ainsi que toutes les entités qui s’approchent de cette position. La seconde requête permet de récupérer l’ensemble des entités situées dans une sphère de rayon de 500 unités autour de la position du guide virtuel.

Chaque contrainte comporte deux aspects :

- ▷ un filtrage des informations présentes dans la représentation courante de l’environnement ;

- ▷ un contrôle du processus de construction et de maintien de la représentation de l'environnement par modification des intérêts associés aux différents comportements de perception.

Ces deux aspects sont déterminés lors de la définition d'une contrainte au travers de deux fonctions : une fonction de *filtrage* et une fonction de *contrôle*. La fonction de filtrage détermine un prédicat à appliquer à une entité visuelle pour valider la contrainte définie. La fonction de contrôle quant à elle définit les modifications d'intérêt à appliquer aux comportements de perception pour favoriser la détection et le maintien en mémoire des entités qui valident la contrainte en question.

Les deux algorithmes suivants présentent ces deux aspects pour la contrainte **type** présentée dans le listing 5.1.

Algorithme 5.4.2: TYPE-FILTER(*visualMemory*, *type*)

– **visualMemory:** contenu courant de la mémoire visuelle
 – **type:** type des entités recherché

local *matchingEntities* $\leftarrow \emptyset$
for each *entity* \in *visualMemory*
 do $\left\{ \begin{array}{l} \text{/* Si le type de l'entité courant est celui recherché */} \\ \text{/* alors celle-ci est ajoutée à la liste des résultats */} \\ \text{if } \textit{entity.type} == \textit{type} \\ \quad \text{then } \textit{matchingEntities} \leftarrow \textit{matchingEntities} + \textit{entity} \end{array} \right.$
return (*matchingEntities*)

La fonction de filtrage effectue un simple test sur la propriété **type** de chaque entité visuelle présente dans la représentation de l'environnement et retourne l'ensemble des entités dont le type correspond au type demandé.

Algorithme 5.4.3: TYPE-CONTROL(*visualMemory*, *type*)

– **visualMemory:** contenu courant de la mémoire visuelle
 – **type:** type des entités recherché

for each *entity* \in *visualMemory*
 do $\left\{ \begin{array}{l} \text{/* On augmente l'intérêt sur les entités correspondant */} \\ \text{/* au type souhaité */} \\ \text{if } \textit{entity.type} == \textit{type} \\ \quad \text{then INCREASEENTITYINTEREST}(\textit{entity}) \end{array} \right.$
 /* On augmente l'intérêt sur le type recherché */
 INCREASETYPEINTEREST(*type*)

La fonction de contrôle suppose deux types d'objets d'intérêt :

- ▷ les entités elles-mêmes ;
- ▷ les différents types identifiables.

Lorsque la contrainte est exécutée, la fonction de contrôle présentée va porter l'intérêt du système de perception visuelle sur les entités dont le type correspond au type demandé ainsi que sur l'ensemble des traitements permettant d'identifier ce type d'entités.

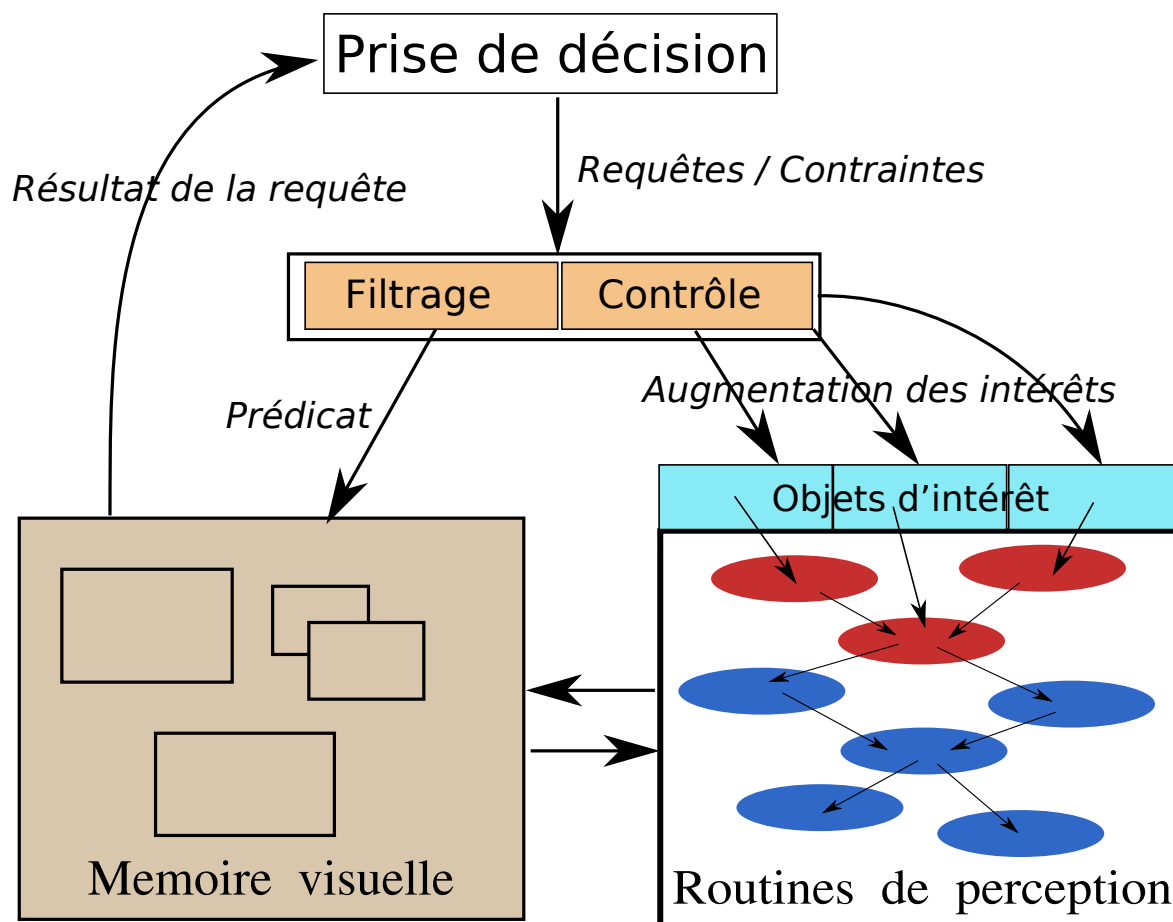


FIG. 5.11 – Vue schématique du mécanisme de contrôle par les requêtes sur la représentation de l'environnement.

La figure 5.11 présente une vue d'ensemble du mécanisme de contrôle par requête que nous proposons. Le processus de prise de décision va interroger la représentation de l'environnement à l'aide d'une requête définissant des contraintes sur les entités qui l'intéresse. Une requête est traduite selon ses deux aspects (contrôle et filtrage). Le filtrage va s'appliquer au contenu courant de la mémoire visuelle et retourner un résultat immédiat. Le contrôle va quant à lui influencer différents objets d'intérêt. Cette influence va avoir un impact sur les traitements effectués en modifiant les stratégies mises en œuvre par les comportements de perception existants.

5.4.5 Précisions sur l'implémentation

Le système de perception visuelle que nous proposons dans ce chapitre est avant tout un système de vision par ordinateur. Il traite des images provenant d'un flux vidéo afin d'en extraire des informations qui permettront de construire une représentation de l'environnement. Ces traitements sont organisés en un ensemble de routines contrôlées par des comportements de perception. Il nous semble important d'apporter ici quelques précisions quant à l'architecture logicielle qui supporte cette proposition. Les détails de cette architecture sont présentés dans l'annexe C et nous ne présentons ici que les aspects qui nous semblent importants vis à vis du fonctionnement de ce système.

L'aspect que nous avons souhaité mettre en avant au sein de cette architecture logicielle est la modularité. L'architecture sur laquelle repose notre implémentation se base sur la notion d'*objet actif*. Un tel objet représente une unité de calcul possédant des capacités de communication (envoi et réception de messages). Nous voyons ici le système de perception visuel comme l'ensemble de ces objets existants à un instant donné qui coopèrent dans le but de construire et mettre à jour la représentation de l'environnement. Chaque objet est spécialisé dans un certain type de tâche comme par exemple le traitement d'image, la classification, le suivi, *etc.*

Les différents objets actifs existants sont organisés sous la forme de processus qui s'exécutent de façon asynchrone. Tous les objets appartenant au même processus sont exécutés de manière séquentielle. L'ordre d'exécution est quant à lui choisi aléatoirement au début de chaque boucle. Une boucle correspond au déclenchement de tous les objets actifs appartenant à un même processus. On peut distinguer deux types d'objets actifs :

- ▷ les objets réactifs qui ne sont exécutés que sur réception d'un message ;
- ▷ les objets proactifs qui sont exécutés lors de chaque boucle.

Ceci a un impact direct sur les possibilités d'exécution des différents composants et le fonctionnement du module de perception. Cela implique notamment l'absence d'une mémoire centralisée au profit d'une approche distribuée.

5.4.6 Bilan

Dans cette section nous nous sommes intéressés aux aspects exécution et contrôle du système de perception visuelle. Nous avons tout d'abord présenté la notion de stratégie de perception qui consiste en le choix et le déclenchement de traitements par des routines de perception particulières caractérisées par leur aspect proactif.

Ces routines particulières, que nous avons désignées sous le terme de comportement de perception, sont influencées par le processus de prise de décision à l'aide de la notion d'intérêt. Cette notion permet de matérialiser l'attention du système de perception en l'obligeant à se focaliser sur des objets d'intérêt particuliers. Les comportements de perception se basent sur

cette notion d'intérêt ainsi que sur le contenu de la représentation de l'environnement afin de sélectionner les traitements devant être exécutés. Un mécanisme d'atténuation tend à annuler au cours du temps les effets de l'influence du processus de prise de décision en équilibrant l'intérêt sur les différents objets qui y sont assujettis.

Nous avons ensuite présenté les mécanismes permettant de contrôler la notion d'intérêt au travers des requêtes effectuées par le processus de prise de décision sur la représentation de l'environnement.

Enfin nous avons mis en avant quelques aspects qui nous paraissent importants au sein de l'architecture logicielle qui supporte l'implémentation du système de perception visuelle que nous proposons parmi lesquels l'exécution asynchrone et la distribution au sein d'objets actifs capables de communiquer entre eux.

L'architecture présentée jusqu'ici reste générique. Dans la section suivante nous nous attachons à la mise en œuvre de ce système dans le but de construire une représentation de l'environnement exploitable par le guide au sens où celle-ci peut lui servir de support pour la mise en place d'une visite guidée.

5.5 Percevoir pour décrire l'environnement

5.5.1 Introduction

Jusqu'ici notre présentation du système de perception est restée relativement abstraite. Dans le cadre de ce manuscrit, l'objectif est de fournir au processus de prise de décision une représentation exploitable de son environnement afin de lui permettre de sélectionner les explications à présenter au visiteur.

Pour être en mesure de construire une représentation utilisable par le processus de décision, il est nécessaire que le système de perception soit capable de (cf. 5.2) :

- ▷ détecter les entités visibles dans l'environnement ;
- ▷ suivre l'évolution de ces entités au cours du temps (position, trajectoire, *etc*) ;
- ▷ identifier les entités suivies parmi un certain nombre de classes connues.

Dans cette section nous présentons la mise en œuvre du système de perception afin de le doter de ces fonctionnalités en utilisant comme cadre l'architecture présentée dans la section précédente. Ceci nécessite de s'intéresser aux deux aspects que nous avons développés précédemment :

- ▷ l'organisation des traitements : quels comportements/routines de perception, quels traitements image, *etc* ;

- ▷ les stratégies mises en œuvre par chacun des comportements, le contrôle effectué par le processus de prise de décision (contraintes sur la représentation de l'environnement).

Il est important de noter que nous présentons ici une mise en œuvre qui peut être considérée comme minimale du système de perception dans le cadre de la description d'un environnement réel tel que celui auquel nous nous intéressons (description des entités autonomes peuplant l'environnement). Il est bien entendu envisageable d'étendre ce système en intégrant d'autres fonctionnalités comme par exemple la reconnaissance d'actions ou la détections de groupes d'entités possédants des propriétés communes.

5.5.2 Organisation pour la description d'un environnement dynamique

Le système de perception que nous proposons repose sur l'utilisation d'un comportement de perception pour chacun des aspects nécessaires à la construction d'une représentation exploitable de l'environnement :

- ▷ le comportement de vigilance qui a pour objectif la détection de nouvelles entités dans le champ visuel. C'est lui qui va créer les entités visuelles en mémoire en calculant leurs propriétés spatiales ;
- ▷ le comportement de suivi dont le but est la mise à jour des propriétés spatiales des entités existantes en mémoire ;
- ▷ le comportement de reconnaissance qui se charge d'identifier les entités contenues dans la mémoire visuelle suivant un ensemble de classes existantes.

Chaque comportement utilise d'autres routines de perception ou routines visuelles dont il se charge d'organiser l'exécution au travers de stratégies de perception. Il se base pour cela sur le contenu de la mémoire visuelle ainsi que sur les objets d'intérêt qu'il propose au processus de prise de décision. La figure 5.12 présente une vue d'ensemble de l'organisation du système de perception visuelle dans le cadre du guide virtuel.

Informations images et routines de perception

Le comportement de vigilance coordonne l'exécution de traitements d'images permettant la détection des entités visuelles et le calcul de leur étendue spatiale. Pour cela il utilise des cartes de propriétés permettant de mettre en valeur cette étendue spatiale (le mouvement par exemple, cf. section 6.2) et sur une routine visuelle permettant l'extraction de ces informations sous la forme de coordonnées d'une boîte englobante dans l'espace image (algorithme de détection de *blobs* et composante connexe).

Le comportement de suivi pour sa part coordonne l'exécution de routines de perception capables de mettre à jour les propriétés spatiales, géométriques et cinématique d'une entité

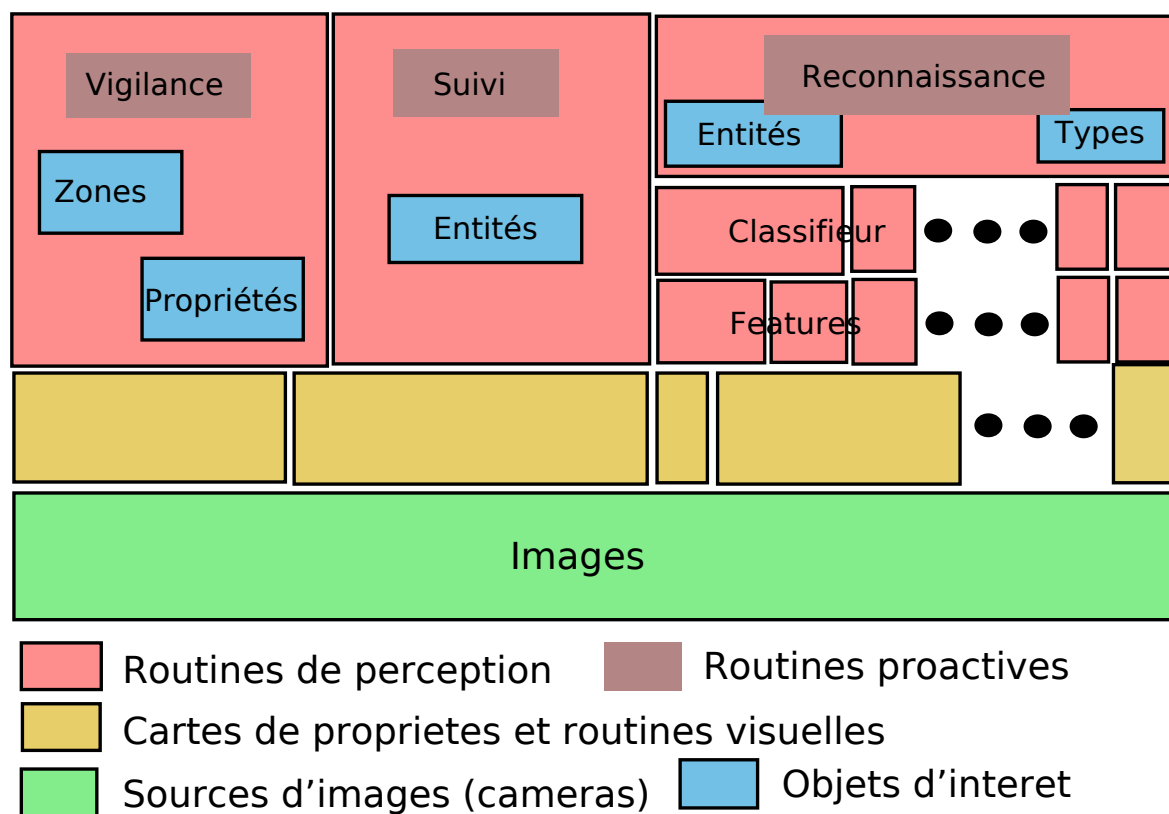


FIG. 5.12 – Organisation du système de perception visuelle dans le cadre du guide virtuel.

particulière. Il utilise pour cela les mêmes routines visuelles et cartes de propriétés que le comportement de vigilance. Ce comportement ne se contente pas de gérer l'exécution des routines et/ou traitements de niveau inférieur. Il effectue des traitements permettant de calculer la trajectoire et la vitesse des différentes entités présentes en mémoire. Il fait également appel à un mécanisme de prédiction basé sur l'utilisation d'un filtre de Kalman [Welch 01] afin de déterminer les zones de l'image où chercher l'entité à mettre à jour. C'est dans ces zones prédites que sont appliqués les traitements bas niveau.

Le comportement de reconnaissance présente la particularité par rapport aux autres comportements de reposer sur plusieurs niveaux de traitement afin de pouvoir obtenir un résultat. Il utilise de façon directe des routines de perception spécialisées dans la classification d'une zone de l'image suivant un ensemble de classes définies. Ces *routines de classification* reposent elles-mêmes sur l'utilisation de plusieurs autres routines afin d'obtenir les informations nécessaires au mécanisme de classification. Chaque routine de classification repose sur l'utilisation d'un vecteur de données pour ses traitements. Ce vecteur est obtenu par concaténation des résultats provenant de routines de perception spécialisées dans le calcul d'informations spécifiques (géométriques, visuelles, spatiales, *etc*). Ces routines font à leur tour appel à différentes cartes de propriétés (masque de mouvement, *etc*) et routines visuelles (extraction de contour, *etc*) afin de fournir un résultat. Ce mécanisme est présenté dans la figure 5.13.

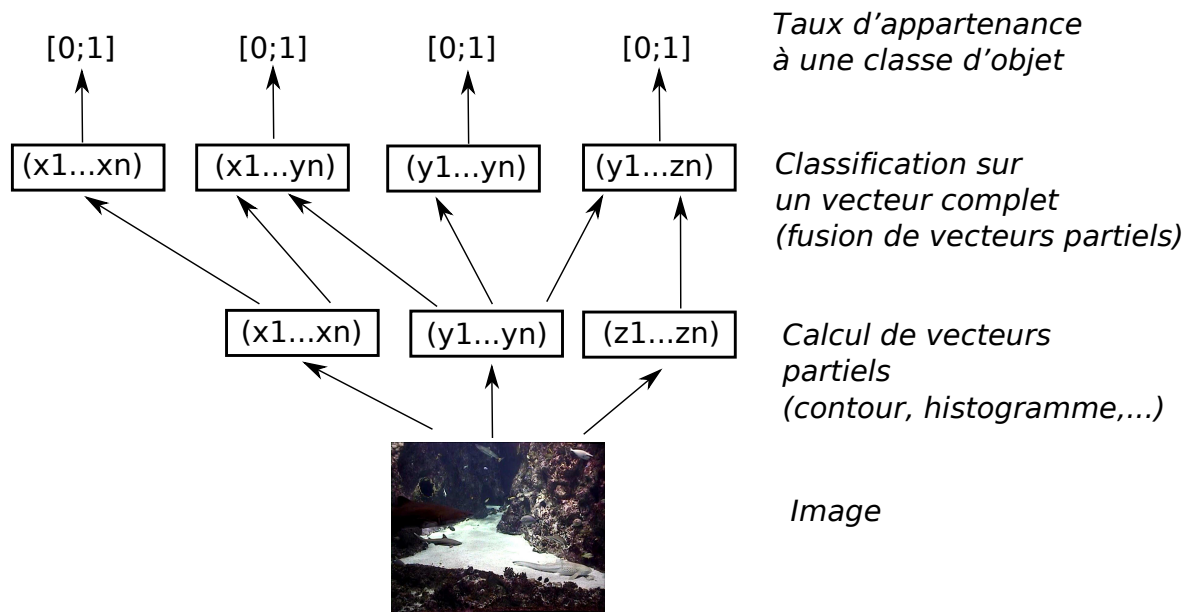


FIG. 5.13 – Les différentes « couches » du mécanisme de reconnaissance.

Objets d'intérêt

Pour chaque comportement, différents objets d'intérêt sont définis en fonction des stratégies de perception qu'ils mettent en œuvre (cf. section 5.5.3).

Le comportement de vigilance propose les objets d'intérêt suivants :

- ▷ des zones du champ visuel définies dans l'espace image ;
- ▷ des propriétés visuelles qui permettent la détection des objets (mouvement par exemple).

Le comportement de suivi propose une gestion d'intérêt basée sur les entités visuelles.

Le comportement de reconnaissance propose la notion d'intérêt pour les deux aspects complémentaire du problème qu'il traite :

- ▷ les entités visuelles pour l'identification d'une entité particulière ;
- ▷ les types pour la recherche d'une entité d'un type particulier parmi les entités existantes.

Le premier point de vue est relativement classique et constitue le fonctionnement habituel du processus de reconnaissance dans les systèmes de vision. Le second point de vue est plus inhabituel mais trouve sa justification dans le fonctionnement et l'objectif du guide virtuel. Celui-ci s'intéresse généralement à un type spécifique d'entité plutôt qu'à une entité particulière. Le comportement de reconnaissance gère ces deux points de vue de façon unifiée grâce à son organisation particulière des routines de classification.

5.5.3 Stratégies de perception

Les comportements mettent en œuvre différentes stratégies de perception en fonction de leurs besoins en terme de résultats et de temps de calcul. Ces stratégies reposent sur l'organisation des traitements présentés précédemment à partir du contenu courant de la représentation de l'environnement et des intérêts modulés par le processus de prise de décision.

Vigilance

Le comportement de vigilance organise l'exécution de traitements de détection suivant deux aspects. Dans le temps, en se chargeant de la gestion du délai entre deux exécutions d'un même traitement. Dans l'espace, en restreignant l'extraction d'informations à partir de la carte de propriétés à une zone spécifique du champ visuel (les côtés par exemple qui sont susceptibles d'être des zones d'apparition d'une nouvelle entité). Pour cela, chaque traitement de détection est décrit par :

- ▷ une expression définissant le calcul d'une carte de propriété ;
- ▷ une zone d'application dans l'image ;
- ▷ le délai minimum entre deux exécutions du traitement ;
- ▷ le délai maximum entre deux exécutions du traitement.

L'exécution d'un traitement correspond ensuite au calcul de la carte de propriété décrite par l'expression puis à l'application d'une routine visuelle permettant d'extraire les différents objets de cette carte de propriétés. Le choix du traitement devant être déclenché s'effectue en fonction de ses paramètres temporels et d'une valeur d'intérêt calculée à partir des objets d'intérêt gérés par le comportement. Un seul traitement au maximum peut être déclenché lors de chaque exécution du comportement et plusieurs traitements peuvent s'exécuter simultanément au sein du système de perception⁴. A chaque pas d'exécution du comportement, certains traitements sont écartés du processus de sélection :

- ▷ les traitements ayant été exécutés depuis un temps inférieur à leur délai minimum d'exécution ;
- ▷ les traitements déjà en cours d'exécution.

Pour les traitements restants, une valeur d'intérêt globale normalisée est calculée suivant la formule suivante :

$$globalInterest = \frac{areaInterest + propertyInterest}{2 * maximumInterest}$$

Les traitements non écartés sont ensuite triés suivant cet intérêt global dans l'ordre décroissant puis la sélection du traitement à exécuter se fait selon l'algorithme 5.5.1.

⁴ Les traitements peuvent nécessiter plusieurs boucles d'exécution des objets actifs.

Algorithme 5.5.1: SELECTVIGILANCEROUTINE(*allRoutines*, *nbEntities*)

– *allRoutines*: ensemble des routines de détection possibles
– *nbEntities*: nombre total d'entités en mémoire

for each *o* \in *allRoutines*

do	<pre> /* Calcul du délai souhaité pour la routine */ local currentInterest \leftarrow o.globalInterest local oneMinusInterest \leftarrow 1 – currentInterest local timeInterval \leftarrow (o.maxTime – o.minTime) local objectTime \leftarrow o.minTime + timeInterval * oneMinusInterest /* Si le temps souhaité est supérieur au temps écoulé */ /* la routine est sélectionnée pour être exécutée */ if objectTime > elapsedTime then return (o) </pre>
-----------	--

Le premier traitement (dans l'ordre d'intérêt global décroissant) dont le délai d'exécution souhaité est supérieur au temps écoulé depuis la dernière exécution du comportement est sélectionné pour être exécuté. Le calcul d'un délai d'exécution souhaité pour chacune des routines de détection existante est fondé sur la formule suivante :

$$t = \text{minDelay} + (\text{maxDelay} - \text{minDelay}) * (1 - \text{interest}), \text{interest} \in [0; 1]$$

Cette formule permet de modifier le temps entre deux exécutions d'une même routine en prenant en compte les temps maximum, minimum et l'influence courante du module pédagogique (figure 5.14). Les valeurs d'intérêt s'équilibrent autour d'une valeur moyenne située à 0.5, situation dans laquelle le délai courant obtenu correspond à la valeur médiane entre le temps minimum et le temps maximum. Plus l'intérêt augmente plus le délai courant obtenu est proche du temps minimum autorisé ; à l'inverse lorsque l'intérêt diminue le délai courant obtenu tend à se rapprocher du temps maximum imparti. Les valeurs d'intérêt ayant tendance à s'équilibrer de façon à maintenir un quantité totale constante (cf. 5.4.3) l'augmentation de l'intérêt pour un type de détection particulier va de fait diminuer l'intérêt pour les autres types de détection, diminuant ainsi le temps entre deux exécutions de la routine de détection intéressante et augmentant le temps entre deux exécutions des autres routines. Sur la figure 5.14, les courbes présentent la variation du délai courant de différentes routines de détection pour plusieurs jeux de paramètres temporels. La section 6.4.3 présente le résultat d'expérimentation sur ce comportement de vigilance et notamment la variation du temps d'exécution entre deux routines en fonction de requêtes du processus de prise de décision.

Ce processus de sélection permet de répartir l'exécution des routines de détection dans le temps et l'espace en fonction de paramètres intrinsèques (les différents délais associés à

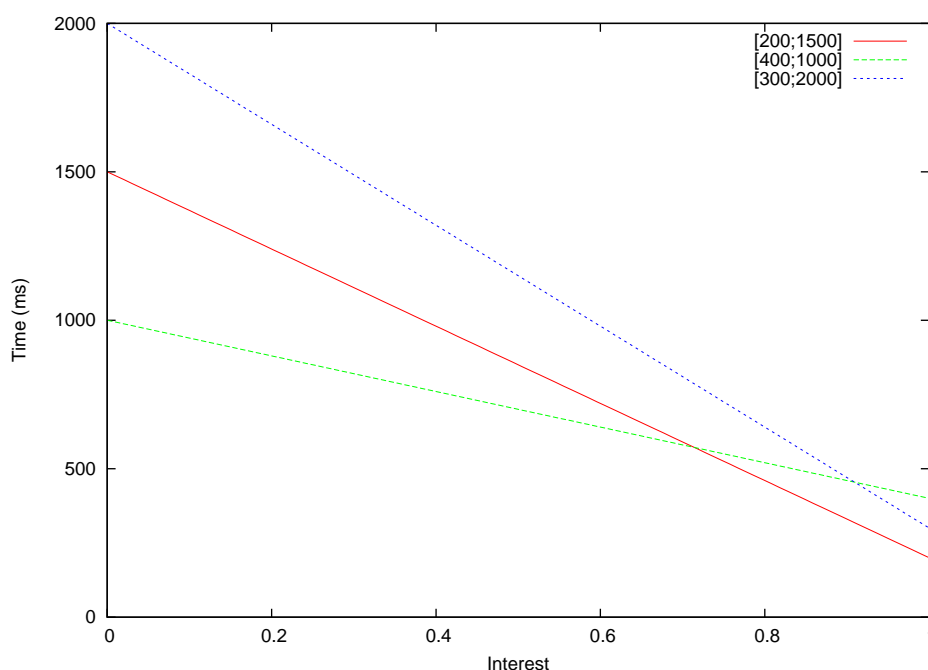


FIG. 5.14 – Exemples de variations du délai courant des routines de détection en fonction des valeurs d'intérêt.

un objet d'intérêt) et de l'influence du module pédagogique (la valeur d'intérêt associée aux différentes zones et propriétés).

Suivi

Le comportement de suivi pour sa part gère l'ordre dans lequel les traitements qu'il coordonne sont déclenchés. La particularité de ces traitements réside dans le fait que ceux-ci doivent être effectués le plus souvent possible afin de maintenir la cohérence de la représentation avec l'environnement réel (figure 5.15). Par conséquent à chaque pas d'exécution, le comportement de suivi a la possibilité de déclencher l'exécution de plusieurs routines de perception, leur exécution étant ensuite conditionnée par le module de traitement d'images et les contraintes en terme de vitesse d'exécution qui lui sont imposées. En effet, étant donnée une puissance de calcul limitée, tous les traitements demandés par les différentes routines ne peuvent être effectués à chaque instant. Le facteur limitant pour l'exécution de ces traitements est ici une limite temporelle, l'exécution des différents traitements ne devant pas excéder une durée maximum définie par avance. La gestion de cette durée est effectuée par le composant responsable du traitement d'image au sein de l'architecture logicielle du système de perception visuelle et dépend de la façon dont l'application est déployée⁵. Du point de vue de l'exécution de traitements par le comportement de suivi, le nombre de routines pouvant être déclenchées n'est pas limité. Les routines ne pouvant être exécutées dans le temps imparti sont considérées

⁵ Ce mécanisme et son influence sont présentés dans la section 6.4. Les détails de l'architecture logicielle sont présentés dans l'annexe C.

comme ayant échouées et le comportement de suivi reçoit une notification de cet échec.

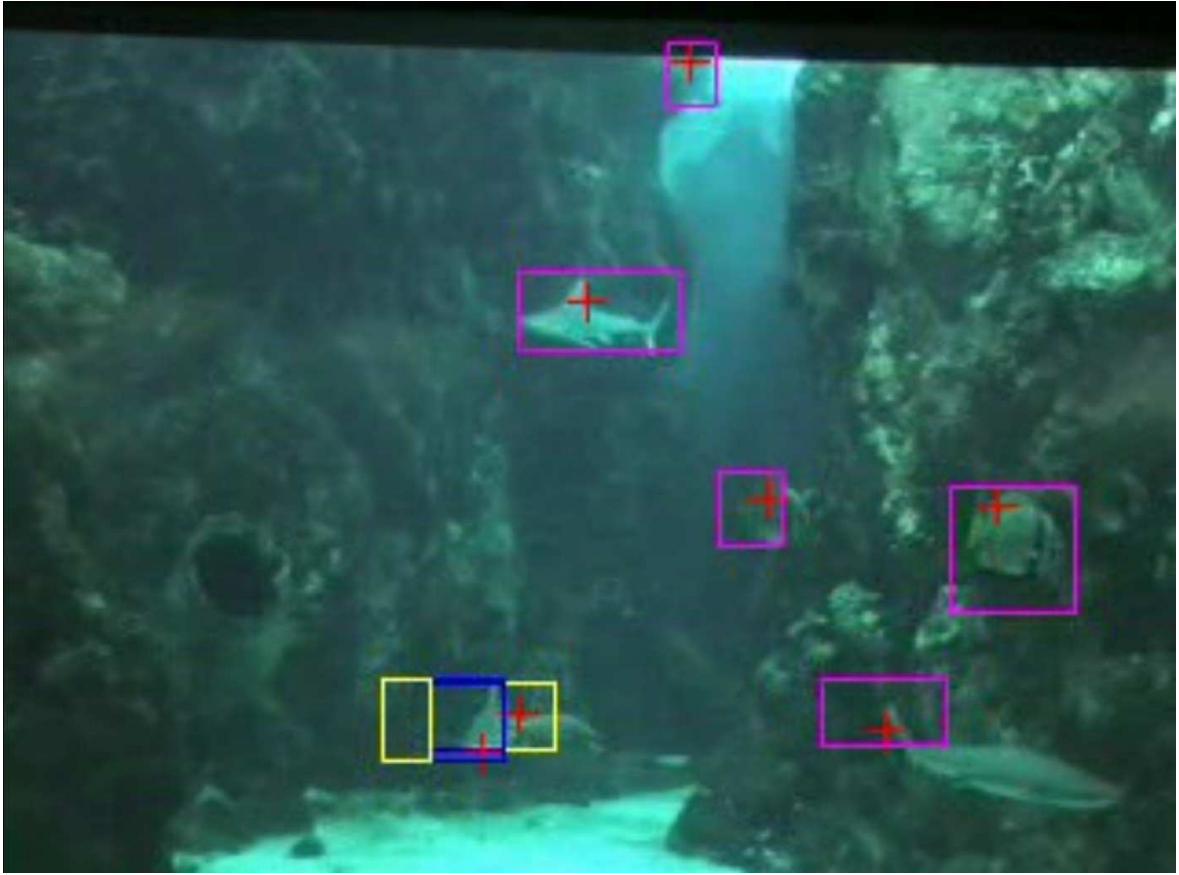


FIG. 5.15 – Capture d'écran du suivi de cibles au sein de notre application à un aquarium marin.

Les traitements dont l'exécution est déclenchée en premier ont donc plus de chances de pouvoir être exécutés dans le temps imparti. C'est sur ce constat que repose la stratégie de perception mise en œuvre dans le comportement de suivi. Celle-ci détermine l'ordre dans lequel les routines doivent être déclenchées, les traitements ne pouvant être effectués dans le temps imparti entraînant à terme la disparition de l'entité de la mémoire visuelle du fait du mécanisme d'oubli mis en place. Il est important de noter que le fait de ne pas respecter la contrainte temporelle n'est pas une solution envisageable. Un tel comportement entraînerait à terme la disparition de la totalité des entités présentes en mémoire du fait d'un décalage croissant entre environnement et représentation.

Lors de chaque pas d'exécution, le comportement de suivi demande l'exécution d'une routine pour chacune des entités existantes dans l'environnement. Cette exécution s'effectue suivant un ordonnancement défini par une valeur de priorité calculée de la façon suivante :

$$priority = interest + \frac{entitySpeed}{maximumSpeed}$$

Ce calcul de priorité permet d'ordonner les objets d'intérêt en tenant compte à la fois

de l'intérêt provenant du module pédagogique et de la vitesse moyenne de l'entité associée à l'objet d'intérêt, les entités se déplaçant plus rapidement nécessitant des mises à jour plus fréquentes de leurs propriétés spatiales.

Reconnaissance

Le comportement de reconnaissance possède une spécificité dans les informations qu'il nécessite afin de déterminer le type des entités visuelles présentes en mémoire. Celle-ci réside dans le fait que plusieurs routines successives sont généralement nécessaires à l'obtention d'un résultat correct. Du fait de cette spécificité, ce comportement met en œuvre des stratégies de perception plus complexes que les comportements précédents.

Pour mettre en œuvre sa stratégie de perception, le comportement de reconnaissance organise les routines de classification dont il dépend en différents *processus de classification*. Un processus regroupe toutes les routines de classification nécessaires à l'identification d'un type particulier d'entité visuelle et cela suivant un ordre déterminé. Au sein de cet ordonnancement les différentes routines de classification sont organisées suivant une complexité croissante (figure 5.16), les routines se trouvant au début d'un processus de classification nécessitant moins de calculs que celles se trouvant à la fin de ce même processus⁶. L'identification d'une entité visuelle est terminée lorsque toutes les routines de classification de l'un des processus de classification ont répondu positivement.

Cette organisation en terme de processus de classification présente plusieurs avantages :

limitation des traitements les traitements nécessaires à l'identification d'un type particulier de façon « certaine »⁷ sont coûteux en terme de temps de calcul et des traitements simples permettent généralement d'écarter certains types pour une entité donnée. La séparation de ces traitements sous la forme de différents processus comportant des traitements de complexité croissante nous permet ici d'écarter certains traitements que l'on sait inutiles pour une entité particulière ;

partage des résultats certains traitements peuvent être communs à l'identification de différents types, les résultats de ces traitements peuvent ici être partagés de façon efficace grâce à la mémoire visuelle ;

points de vue multiples cette organisation nous permet de gérer les deux points de vue de classification et autorise à la fois l'identification d'une entité donnée et la recherche d'entités d'un type spécifique parmi toutes les entités encore non identifiées ;

flexibilité une telle organisation permet de plus une grande souplesse dans l'ajout ou la modification des types que le système de perception visuelle est capable d'identifier contrairement à des approches plus « classiques » se basant par exemple sur l'utilisation d'un classifieur unique pour l'ensemble des types existants.

⁶ Cet ordonnancement des routines de classification est effectué de façon statique lors de la configuration du système.

⁷ Dans le cadre de nos travaux nous cherchons à nous assurer que le type associé à une entité visuelle est exact. Il serait en effet désastreux du point de vue des explications du guide de présenter des informations sur un type d'entité donné en associant ces explications à une entité identifiée de façon incorrecte.

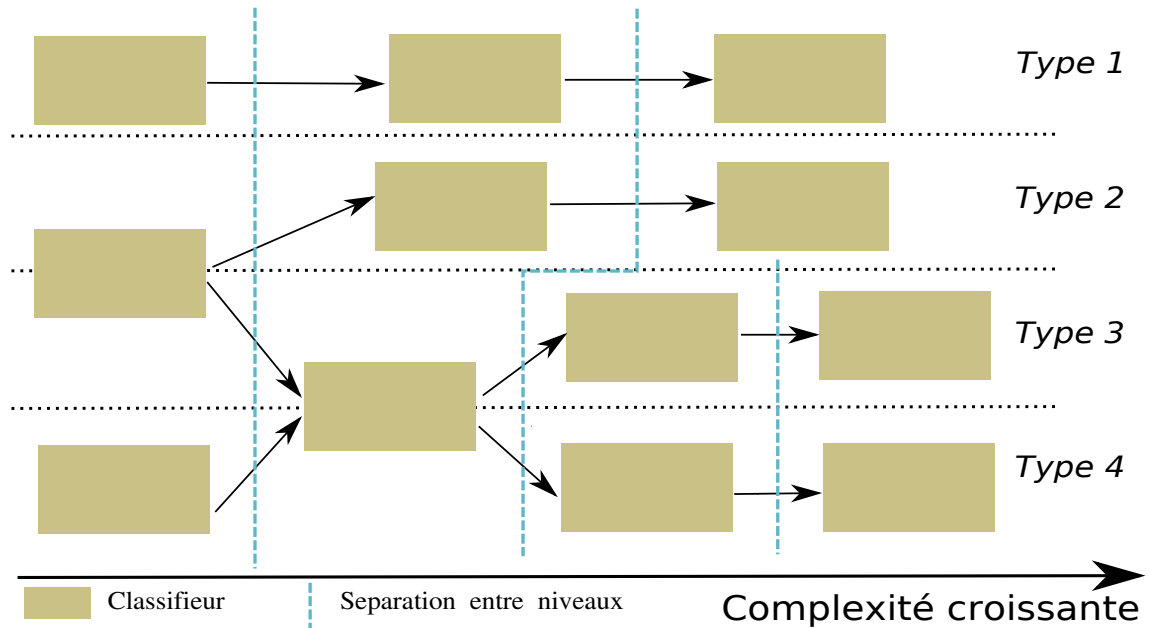


FIG. 5.16 – Organisation des routines pour le comportement de reconnaissance en processus de classification.

Pour chaque entité visuelle, le comportement de reconnaissance conserve en interne l'état d'avancement de son identification au sein de chacun des processus de classification. Une entité visuelle est considérée comme identifiée lorsque l'un des processus de classification a atteint son terme, c'est à dire que la dernière routine de classification qui le constitue a renvoyé un résultat positif. Si l'une des routines d'un processus de classification renvoie un résultat négatif, celui-ci est alors considéré comme invalidé et ne sera pas testé plus avant.

A chaque exécution du comportement de reconnaissance, une seule routine de classification est exécutée pour une entité visuelle donnée. Pour un même processus de classification, le résultat de l'exécution d'une routine donnée est attendu avant de déclencher l'exécution de la routine suivante au sein de ce processus.

Le choix de la routine à déclencher pour une entité visuelle à un pas d'exécution donné se fait au sein de l'ensemble des routines disponibles pour cette entité, c'est à dire les routines pour lesquelles les routines précédentes au sein d'un processus de classification ont été exécutées et ont retourné un résultat positif. Ce choix se fait également uniquement sur les entités visuelles encore non identifiées jusqu'à présent pour des raisons évidentes. L'algorithme 5.5.2 présente le processus de sélection des routines de classification déclenchée par le comportement de reconnaissance lors de chaque pas d'exécution. Le résultat de cet algorithme consiste en un couple (*entity*; *type*) qui sera déclenché. Si plusieurs solutions sont possibles l'une d'elles est sélectionnée de manière aléatoire.

Cet algorithme permet de prendre en compte à la fois l'intérêt du module pédagogique pour une entité ou un type particulier et l'état d'avancement d'un processus de classification pour une entité. Ceci est effectué à l'aide de la formule de calcul de priorité suivante pour une entité et un processus de classification donnés :

Algorithme 5.5.2: SELECTCLASSIFICATIONROUTINE(*types*, *entities*)

```

– types: processus de classification
– entities: entités non identifiées présentes en mémoire
selectedRoutines ← ∅
maxInterest ← −1
/* Calcul de l'ensemble des couples (entity; type) */
/* d'intérêt total maximum */
for each entity ∈ entities
do {
  for each type ∈ types
  do {
    /* Calcul de l'intérêt total associé au couple (entity; type) */
    currentInterest ←  $\frac{(type.interest + entity.interest) * (entity.currentLevel(type) + 1)}{type.nbLevels}$ 
    /* Si cet intérêt est le plus important jusqu'à présent */
    /* on remet à zéro la liste des couples sélectionnés */
    /* Sinon on ajoute le couple courant à la liste des couples */
    /* sélectionnés */
    if currentInterest > maxInterest
    then { maxInterest ← currentInterest
          selectedRoutines ← ∅
        }
    else if currentInterest = maxInterest
    then selectedRoutines ← selectedRoutines + (entity; type)
  }
}
/* On choisit un couple aléatoirement parmi l'ensemble des couples */
/* d'intérêt maximum */
return (RANDOM(selectedRoutines))

```

$$priority = (typeInterest + entityInterest) * \frac{currentLevel}{nbLevels}$$

Ainsi la priorité d'un couple (*entity; type*) correspond à la somme des intérêt de chacun des éléments du couple pondérée par l'état d'avancement pour l'entité dans le processus de classification en question (figure 5.17).

5.5.4 Contrôle du processus de prise de décision

Partant des différents objets d'intérêt définis par les comportements de perception que nous venons de présenter, nous nous intéressons maintenant aux différentes contraintes que nous proposons pour l'interrogation de la représentation de l'environnement et à la façon dont celles-ci vont influencer l'intérêt des comportements de perception.

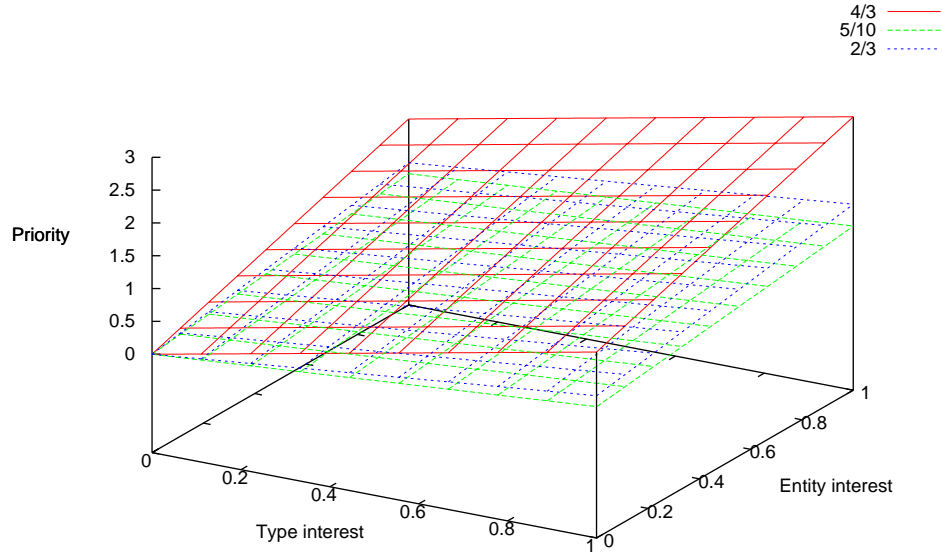


FIG. 5.17 – Variations de la priorité des routines de reconnaissance en fonction de l'intérêt porté sur les entités et les types. Chaque plan correspond à un couple possédant un état d'avancement particulier.

Le processus de prise de décision interroge la représentation de l'environnement en plusieurs occasions :

- ▷ lorsqu'il sélectionne le sujet courant des explications ;
- ▷ lorsqu'il cherche une entité visuelle correspondant au sujet courant afin de s'en servir comme support visuel aux explications qui vont être générées.

Un certain nombre de contraintes sont nécessaires lors de ces différentes occasions.

- ▷ une contrainte sur le type des entités. Cela peut servir dans le cadre d'un expert favorisant un sujet si le type d'entités qui y est éventuellement associé est présent dans la représentation de l'environnement ;
- ▷ une contrainte sur la distance des entités par rapport à la position courante de la représentation du guide virtuel. Cela est utile si l'on veut sélectionner un support visuel dans une certaine zone autour de cette représentation ou s'il l'on cherche à simuler un champ de vision limité lors de la sélection du sujet (favoriser les entités proches par exemple) ;
- ▷ une contrainte sur des zones de l'image ;
- ▷ une contrainte sur l'identifiant d'une entité particulière.

Le premier aspect de ces contraintes est relativement évident et consiste en un filtrage du contenu de la représentation de l'environnement. Dans la suite de cette section nous allons donc mettre l'accent sur le contrôle et l'influence des différents objets d'intérêt.

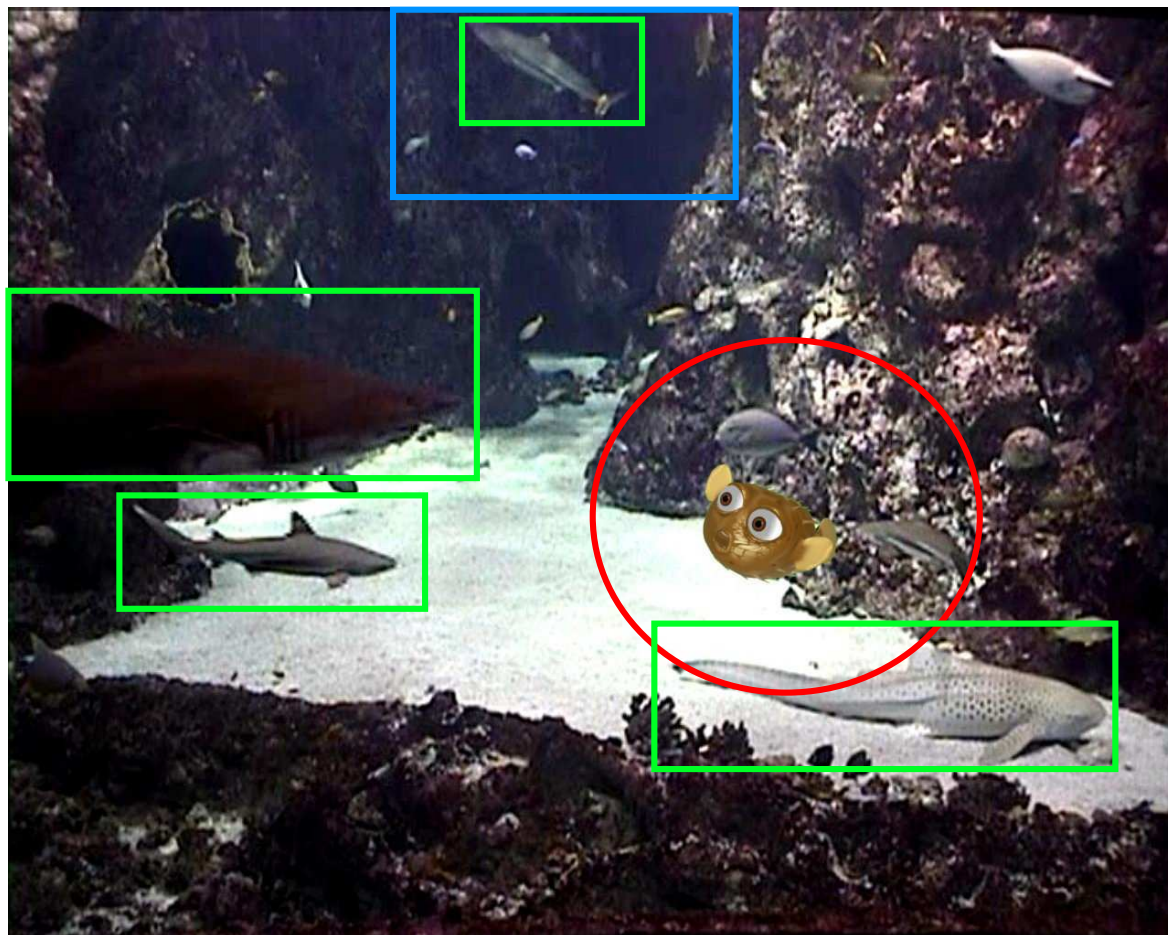


FIG. 5.18 – Représentation graphique des différentes contraintes pouvant être utilisées lors des requêtes sur la représentation de l'environnement.

La contrainte sur le type des entités permet de récupérer l'ensemble des entités dont le type est égal à celui spécifié en paramètre de la contrainte (en vert sur la figure 5.18). Du point de vue du contrôle cette contrainte va influencer tous les comportements selon différents points de vue :

- ▷ le comportement de vigilance va être influencé en faveur de la détection des propriétés et zones qui correspondent à ce type particulier d'entité. Ceci repose sur la définition de connaissances *a priori*⁸ permettant d'associer ces différentes informations ;

⁸ La définition de ces connaissances se fait pour l'instant de façon implicite dans la définition de la partie contrôle de cette requête. Il serait néanmoins important de s'intéresser à la modélisation de ces connaissances de manière plus formelle.

- ▷ le comportement de suivi va être influencé de telle sorte que toutes les entités qui sont du type recherché seront traitées en priorité ;
- ▷ le comportement de reconnaissance va être influencé afin de se concentrer sur l'identification d'entités du type spécifié parmi les entités encore non reconnues (plutôt que l'identification d'une entité particulière).

La contrainte sur la distance va influencer uniquement le comportement de suivi afin de favoriser les entités qui se trouvent dans la zone entourant la représentation du guide virtuel définie par la distance spécifiée (en rouge sur la figure 5.18). De même la contrainte sur la zone du champ visuel va influencer le suivi pour les entités contenues dans la zone spécifiée (en bleu sur la figure 5.18). Enfin la contrainte sur une entité particulière (par l'identifiant unique qui lui est associé) va influencer le suivi et la reconnaissance de cette entité.

Les différentes routines de perception encapsulent un certain nombre de connaissances relatives au domaine d'application. C'est le cas notamment des routines de classification que nous avons présentées ainsi que des différents comportements. Néanmoins il paraît clair que l'ensemble des connaissances pouvant s'avérer nécessaires ne peuvent être définies au sein de ces routines. Ainsi certaines connaissances doivent être précisées au niveau de la définition des contraintes permettant au processus de prise de décision d'accéder au contenu de la représentation de l'environnement. C'est ici le cas pour l'influence sur le comportement de vigilance et l'aspect contrôle de la contrainte sur le type des entités présentes dans l'environnement.

5.5.5 Bilan

Dans cette section nous avons présenté la mise en œuvre du système de perception que nous proposons dans le cadre du guide virtuel et plus généralement de la description d'un environnement dynamique peuplé d'entités autonomes.

Cette mise en œuvre repose sur trois comportements de perception permettant de gérer l'exécution des traitements :

- ▷ un comportement de vigilance qui met en œuvre une stratégie de détection des entités peuplant l'environnement observé ;
- ▷ un comportement de suivi dont l'objectif est de maintenir à jour les informations spatiales des entités présentes dans la représentation de l'environnement ;
- ▷ un comportement de reconnaissance ayant pour but l'identification des entités présentes en mémoire.

Des objets d'intérêt sont proposés par chaque comportement de perception. Ces objets dépendent des stratégies de perception mise en œuvre par chacun des comportements. Le comportement de vigilance gère l'intérêt de zones du champ visuel et de propriétés visuelles. Le comportement de suivi gère un intérêt lié aux entités présentes en mémoire. Le comportement

de reconnaissance présente la notion d'intérêt sur les entités présentes en mémoire et les types d'entités qu'il est capable d'identifier.

Les stratégies mises en œuvre par les comportements prennent en compte l'influence du processus de prise de décision au travers des objets d'intérêt qu'ils proposent et l'influence de l'environnement au travers du contenu de la représentation de l'environnement. Ainsi, en plus des différentes valeurs d'intérêt, le comportement de vigilance tient compte du nombre d'entités présentes en mémoire à un instant donné afin d'organiser les routines de détection qu'il coordonne ; le comportement de suivi tend à favoriser les entités visuelles dont la vitesse est plus élevée partant de l'hypothèse que les propriétés spatiales de ces entités sont sujettes à de plus grandes variations que les autres entités ; enfin le comportement de reconnaissance tient compte de l'état d'avancement des processus de classification des différentes entités afin de n'explorer que les traitements *a priori* pertinents.

Nous avons finalement présenté les différentes contraintes que nous avons définies afin de permettre au guide d'interroger la représentation de l'environnement de manière pertinente par rapport à son objectif de description des différentes entités présentes en mémoire. Pour chacune de ces contraintes nous avons mis en avant les mécanismes de contrôle associés envers les différents comportements de perception existants.

5.6 Conclusion

Dans ce chapitre nous avons présenté l'architecture et le fonctionnement du système de perception visuelle permettant au guide virtuel de connaître l'état de l'environnement qu'il cherche à décrire. Nous avons tout d'abord insisté sur la nécessité pour le guide virtuel de disposer d'une représentation de l'état de l'environnement à tout instant du fait des traitements coûteux nécessaires à la perception de ce dernier. Nous avons ensuite présenté la notion de représentation partielle de l'environnement et la nécessité de la mise en place de mécanismes de contrôle permettant sa construction de façon adaptée par rapport aux besoins du processus de prise de décision du guide virtuel.

Ensuite nous avons présenté l'organisation des informations au sein du système de perception. Cette organisation consiste en une approche hiérarchique des traitements centrée autour d'une mémoire visuelle composée d'entités décrites par différentes propriétés. C'est cette mémoire qui sert de représentation de l'environnement pour le processus de prise de décision. Des notions de cartes de propriétés et de routines visuelles permettent l'extraction d'informations à partir des images fournies par différentes sources (des caméras). Un ensemble de routines de perception permettent le calcul des différentes propriétés définissant les entités présentes en mémoire. Pour cela ces routines peuvent reposer sur des traitements au niveau image ou le calcul de propriétés par d'autres routines.

Nous avons alors présenté la façon dont les traitements sont exécutés au sein de cette architecture en introduisant la notion de stratégie de perception. Ces stratégies, mises en œuvre par des routines spécialisées désignées sous le terme de comportements de perception permettent de coordonner le déclenchement et les paramètres des routines de perception de niveau inférieur en tenant compte du contenu courant de la représentation et des besoins

du processus de prise de décision. Ces besoins sont modélisés à l'aide de la notion d'objets d'intérêt définis pour chacun des comportements existants. Nous avons également apporté quelques précisions quant à l'architecture logicielle supportant l'implémentation du système de perception visuelle en mettant en avant ses possibilités de distribution et l'exécution asynchrone des différentes routines visuelles.

Enfin nous avons appliqué ce modèle de perception visuelle au cadre de nos travaux, à savoir la construction d'une représentation d'un environnement dynamique peuplé d'entités autonomes évoluant de manière imprévisible. Pour cela nous avons identifié trois comportements nécessaires à la mise en place d'un processus de représentation de l'environnement exploitable par le processus de prise de décision du guide virtuel :

- ▷ un comportement de vigilance capable de détecter des entités présentes dans le champ visuel ;
- ▷ un comportement de suivi capable de mettre à jour la position et la taille des différentes entités présentes en mémoire ainsi que de calculer leur propriétés cinématiques ;
- ▷ un comportement de reconnaissance capable d'identifier les entités visuelles présentes en mémoire ou de rechercher une entité d'un type donné parmi ces entités.

Ces différents comportements reposent sur l'utilisation de routines visuelles et routines de perception spécifiques dont ils coordonnent l'exécution de façon autonome. La coopération des différents comportements au travers de cette gestion des routines permet d'effectuer les traitements nécessaires au calcul des différentes propriétés des entités visuelles ainsi qu'au maintien en mémoire de ces entités, fournissant ainsi les informations nécessaires au déroulement de la visite guidée. Nous avons alors décrit le fonctionnement de l'influence du processus de prise de décision en détaillant les contraintes utilisables pour l'interrogation de la représentation.

Partie IV

— Application et expérimentations

—

Chapitre 6

Application et expérimentations

6.1 Introduction

Jusqu'ici nous sommes restés relativement abstraits dans la description de notre proposition. Ce chapitre est l'occasion pour nous de présenter des détails quant à l'application de notre solution à un problème réel (la visite guidée d'un aquarium d'Océanopolis). Nous commençons par préciser les contraintes propres à cet aquarium marin avant de détailler la mise en œuvre du guide virtuel autonome dans ce cadre particulier. Nous présentons les deux aspects développés tout au long de ce manuscrit – la description du réel et la perception du réel – et mettons en avant les contraintes inhérentes au cadre applicatif.

C'est également l'occasion de présenter les résultats d'expérimentations menées sur chacun des deux aspects développés tout au long de ce manuscrit. Ces expérimentations, effectuées dans des conditions contrôlées hors du cadre applicatif, nous permettent de mettre en évidence le fonctionnement des différents mécanismes mis en œuvre.

6.2 Application : le projet SIRENE

6.2.1 Besoins en terme de présentation d'informations

Les aquariums marins et Océanopolis en particulier sont comme la plupart des musées à la recherche de nouveaux moyens de communication avec les visiteurs, de nouvelles façons de leur présenter des informations (cf. section 1.1.2).

L'aquarium d'Océanopolis propose à ses visiteurs une cinquantaine de bassins présentant plus de mille espèces animales différentes dans des habitats reconstitués. Les différentes espèces de chaque bassin sont présentées de façon succincte sur les côtés de celui-ci sous la forme de fiches (figure 6.1). Chaque fiche présente l'espèce à l'aide d'une photo, d'un nom scientifique et de renseignements divers sur son habitat ou ses habitudes alimentaires (figure 6.2). L'espace

disponible sur une telle fiche est fortement limité et les informations qu'elle présente ne peuvent être adaptées au public qui les consulte. De plus ces fiches sont généralement physiquement éloignées du bassin de telle sorte que le visiteur se voit contraint de mettre en relation de lui-même les informations présentées sur ces fiches et le contenu de l'aquarium. Ceci nécessite la plupart du temps que le visiteur se déplace entre les informations présentées sur les côtés et le contenu du bassin afin de retrouver l'espèce en question. Le poisson présenté peut d'ailleurs ne pas être visible à cet instant. Le nombre de fiches pouvant être disposées aux alentours des bassins est de plus limité et par conséquent seule une partie des espèces présentes sont expliquées au visiteur par ce moyen.



FIG. 6.1 – Des visiteurs face à l'un des bassins d'Océanopolis.

Étant données ces limitations, le moyen de communication le plus efficace pour présenter le contenu des différents bassins aux visiteurs (entre 450 000 et 500 000 par an) reste la visite guidée. Celle-ci pose néanmoins des problèmes de logistique (organisation des visites, planification) et de moyens. Le nombre de guides disponibles étant limité (entre 15 et 40 suivant les saisons), tous les visiteurs ne peuvent pas en profiter (seulement 10% des visiteurs bénéficient d'une visite guidée).

Nous présentons ici l'application de notre système à l'explication d'un des bassins du pavillon tropical d'Océanopolis. Ce travail s'effectue dans le cadre du projet ANR SIRENE¹ (**S**ystème **I**chtyologique de **R**enseignement **E**ducatif **N**autique**E**). Un tel projet nous offre un cadre d'application adapté pour notre système tandis que l'application elle même présente un intérêt évident pour Océanopolis du point de vue de sa mission culturelle et pédagogique envers les visiteurs.

¹ Ce projet s'effectue en collaboration avec Océanopolis (<http://www.oceanopolis.com>), Virtualys (<http://www.virtualys.fr>) et Polymorph Software (<http://www.polymorph.fr>).

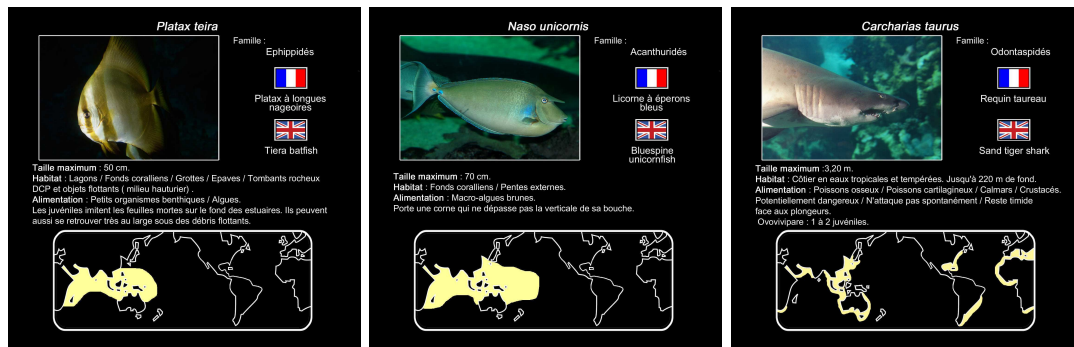


FIG. 6.2 – Exemples de fiches de présentation de différents poissons d'un des bassins du pavillon tropical d'Océanopolis.

6.2.2 Mise en œuvre

Le bassin d'un aquarium marin est peuplé d'un ensemble d'entités autonomes (les poissons) que le guide virtuel doit tenter de décrire aux visiteurs. La vitre du bassin offre un champ de vision limité où ces différentes entités peuvent apparaître ou disparaître à tout moment.

Bien que cela constitue un cadre relativement approprié pour la mise en œuvre du guide virtuel, il présente certaines difficultés non négligeables :

- ▷ pour le rendu et la visualisation : le système final doit être intégré correctement au circuit de visite du musée en respectant la contrainte de point de vue fixe que nous nous sommes imposée ;
- ▷ pour la perception : l'environnement aquatique et la vitre du bassin entraînent des distortions et artefacts visuels (en particulier des caustiques) qui rendent difficiles les traitements d'images utilisés. De plus les entités observées sont relativement difficiles à reconnaître dans ces conditions peu contrôlées.

Le système que nous proposons dans le cadre de notre application est mono-utilisateur et consiste en :

- ▷ un système de visualisation placé dans le circuit de visite ;
- ▷ deux caméras permettant de percevoir le bassin à décrire ;
- ▷ un ensemble d'ordinateurs chargés d'effectuer les traitements nécessaires ainsi que le rendu du personnage incarnant le guide virtuel.

Le système de visualisation repose sur l'utilisation d'une vitre semi-transparente placée devant l'aquarium sur laquelle est projetée la représentation du guide virtuel à l'aide d'un écran LCD. Le visiteur se place devant cette vitre et perçoit le reflet du guide virtuel qui lui apparaît alors immergé dans l'environnement réel (figure 6.3).



FIG. 6.3 – Le bassin du point de vue du visiteur au travers du système de visualisation et l'utilisateur devant le prototype du système.

La perception de cet environnement repose sur l'utilisation des deux caméras mobiles placées face à l'environnement à décrire. Le champ de vision des caméras est adapté de façon à coïncider avec le point de vue du visiteur lorsqu'il utilise le système. Les différents ordinateurs utilisés permettent d'exécuter le système de perception (1 ordinateur) ainsi que le processus de prise de décision et le système de rendu (1 ordinateur). Nous avons réalisé un prototype qui est en cours de finalisation. Celui-ci consiste en un ensemble de données d'explication et de comportements concernant trois grandes espèces de poissons ainsi que le bassin lui-même.

Nous avons également présenté une version modifiée de ce prototype lors de la Nuit des Chercheurs à Océanopolis. Dans le cadre de cette démonstration, un autre système de visualisation a été utilisé pour restituer le guide immergé dans l'environnement réel sur un écran vidéo (figure 6.4). Dans ce cas un *chroma-keyer* permet l'incrustation de l'image du guide virtuel dans l'image de l'aquarium réel.

6.2.3 Comportement et discours

Les poissons qui peuplent le bassin de l'aquarium que le guide cherche à expliquer au visiteur se déplacent en permanence, entrant et sortant du champ de vision des caméras et du visiteur. Les explications du guide doivent donc pouvoir être adaptées à cette contrainte de façon à correspondre au contenu de l'environnement à l'instant où elles sont fournies. Le système que nous proposons est fait pour ce genre de situation en permettant des changements fréquents dans le cours des explications tout en tentant de maintenir une cohérence globale nécessaire à la compréhension de la présentation de l'aquarium. Pour cela il nécessite la définition d'un



FIG. 6.4 – Démonstration dans le cadre de la Nuit des Chercheurs du 26 Septembre 2008.

certain nombre de données et de paramètres de discours ainsi que des différents comportements du guide virtuel. La création de ces données s'est faite en collaboration avec les personnes responsables de la création des contenus pédagogiques à Océanopolis avec la coopération des biologistes responsables du bassin choisi pour l'application.

Comportements

Le fonctionnement du système a été pensé sous la forme de courtes sessions d'explication d'environ cinq minutes, les visites guidées, au cours desquelles le guide peut aborder différents sujets. Ces sujets sont en rapport avec les espèces présentes dans le bassin et traitent de thèmes spécifiques comme la reproduction, la nage ou encore l'habitat de ces espèces. Chaque session d'explication est « encadrée » par des phases d'introduction et de conclusion permettant au guide de se présenter et de notifier la fin de la visite. Entre chaque visite complète (introduction, session d'explication et conclusion) le guide virtuel passe dans une phase d'attente au cours de laquelle il évolue de manière aléatoire dans l'environnement.

La figure 6.5 détaille ce comportement global du guide virtuel (cf. section 3.4). Les carrés

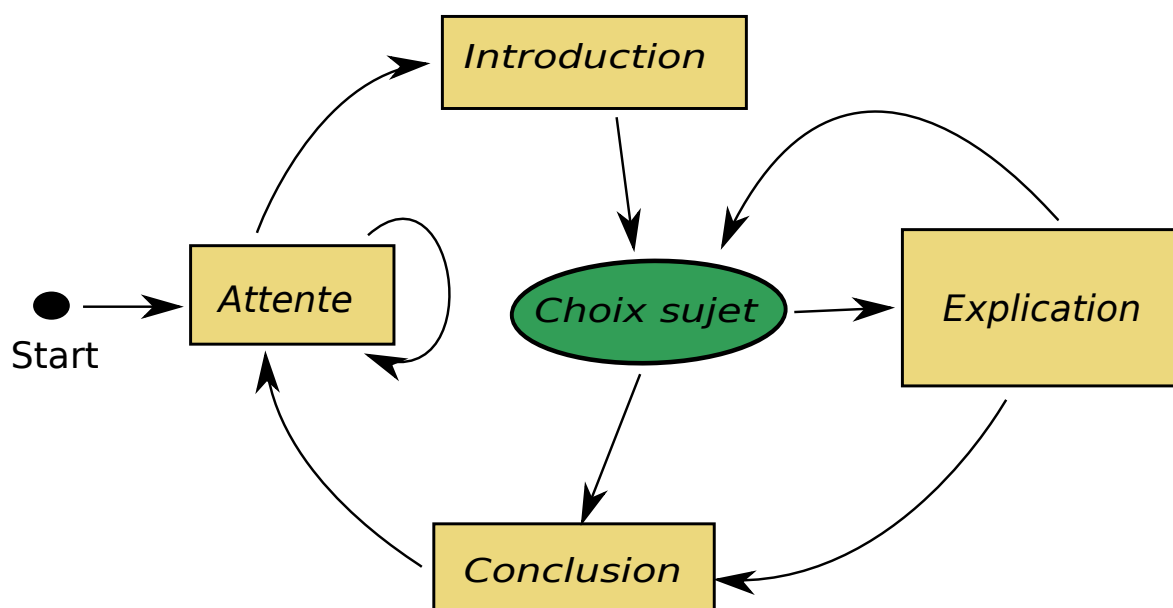


FIG. 6.5 – Comportement global du guide virtuel dans le cadre de l’application.

y représentent des états composites et les ellipses des états atomiques.

L’état *Attente* est un état composite constitué de comportements faisant évoluer le guide de façon crédible dans l’environnement. Dans cet état, les différents comportements locaux sont sélectionnés de manière aléatoire et ce tant qu’un événement signalant le déclenchement de la visite n’est pas détecté. Lorsque le comportement sélectionné est arrivé à son terme, un autre comportement est alors sélectionné de la même façon. Les comportements de cet état présentent tous la particularité d’être interruptibles afin de permettre le déclenchement de la visite à n’importe quel moment. La figure 6.6 présente un exemple de comportement d’attente. Ce comportement permet au guide virtuel de se déplacer d’une position de l’environnement à une autre, cette position étant sélectionnée de façon aléatoire. Ici le guide peut être interrompu lorsqu’il est en cours de déplacement mais pas lorsque le comportement démarre ou qu’il est arrivé à la position choisie.

Une fois cette visite déclenchée, le guide passe dans l’état *Introduction*. Celui-ci est également un état composite et consiste en une séquence de comportements devant être exécutés en totalité avant le début de la session d’explication.

Une fois cette séquence terminée, la session d’explication commence. Celle-ci consiste en une succession de phases de sélection de sujet et de phases d’explication et ce jusqu’à ce que la durée autorisée pour la session soit écoulée ou que la totalité des sujets disponibles aient été expliqués.

La phase de sélection de sujet se décompose de la façon suivante :

- ▷ Faire voter les experts sur l’ensemble des sujets disponibles (sujets non encore abordés et non terminés) ;
- ▷ Choisir le sujet courant aléatoirement parmi les sujets ayant le plus grand nombre de votes ;

▷ Si un sujet est sélectionné :

- ◊ Tenter de sélectionner une entité visuelle appropriée (la plus proche du guide en rapport avec le sujet sélectionné) ;
- ◊ Sélectionner une transition de discours parmi l'ensemble des transitions valides à l'instant courant.

Lorsqu'aucun sujet n'a pu être sélectionné, la phase de conclusion est déclenchée. Il est important de noter ici que la configuration d'experts utilisée n'implique aucun mécanisme de veto assurant ainsi qu'un sujet peut toujours être sélectionné à condition qu'il reste des sujets disponibles (non terminés).

La phase d'explication se base sur le contexte courant de la visite (sujet sélectionné, historique, transition, ...) afin de sélectionner un comportement approprié parmi les comportements suivants :

- ▷ explication directe sans entité visuelle sélectionnée ;
- ▷ explication avec transition sans entité visuelle sélectionnée ;
- ▷ explication directe avec entité visuelle sélectionnée ;
- ▷ explication avec transition avec entité visuelle sélectionnée.

Les comportements sans entité visuelle sélectionnée font se déplacer le guide de manière aléatoire alors qu'il exécute les explications. Les comportements avec entité visuelle vont contrôler les déplacements du guide afin de le placer à côté de l'entité sélectionnée. Les comportements avec transition présentent la particularité d'imposer l'exécution d'un élément de transition avant de démarrer une explication. Ces différents comportements d'explication possèdent une gestion des interruptions similaire, celles-ci étant autorisées à la fin de l'exécution de chaque élément de discours.

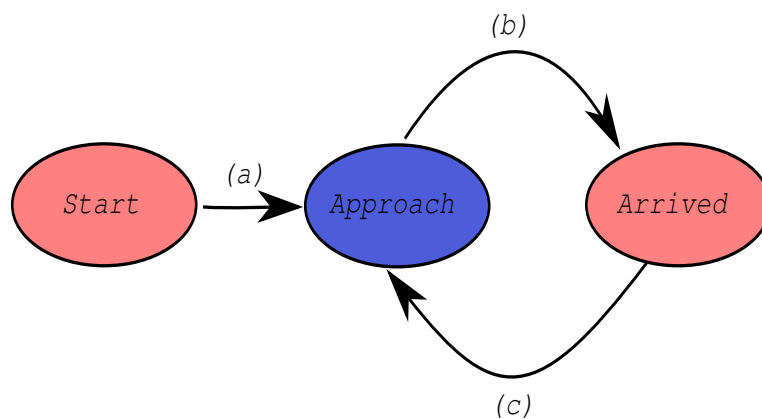


FIG. 6.6 – Exemple de comportement d'attente.

La figure 6.7 présente un exemple de comportement d'explication directe avec entité visuelle. Ici le guide s'approche de l'entité visuelle qu'il cherche à décrire. Un fois suffisamment

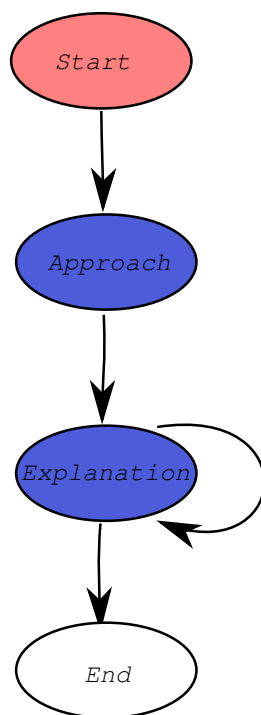


FIG. 6.7 – Comportement d’explication directe avec entité visuelle.

proche, il commence son explication. Il peut être interrompu à tout moment, soit lorsqu’il est en cours d’approche (si une entité visuelle plus intéressante est repérée par exemple) ou pendant son explication à condition que l’élément de discours en cours d’exécution soit terminé. La figure 6.8 présente quant à elle un comportement d’explication incluant une phase de transition. Ici le guide s’approche de l’entité visuelle à décrire puis, une fois arrivé à une distance suffisante, il commence sa transition avant de commencer son explication. On joue ici sur le contrôle de l’interruptibilité pour s’assurer que l’on effectue la transition ainsi qu’un minimum d’explication avant de pouvoir être interrompu dans le but de maintenir la cohérence du discours du guide virtuel.

Discours

Trois espèces de poissons ont été considérées dans le cadre des explications que peut fournir le guide :

- ▷ le *Platax* ;
- ▷ le *Naso* ;
- ▷ les *Requins* et plus particulièrement le *Requin taureau*.

A cela viennent s’ajouter des explications d’ordre général sur le bassin ainsi que des données permettant d’introduire et de conclure la visite. Pour la démonstration utilisée lors de la

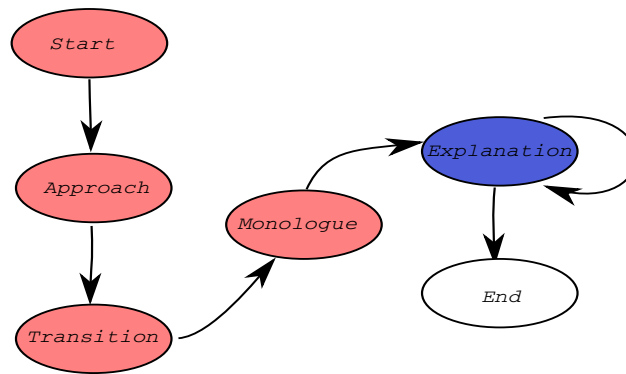


FIG. 6.8 – Comportement d’explication avec entité visuelle comportant une phase de transition.

Nuit des Chercheurs, des données supplémentaires ont été définies et concernent le guide lui-même et son fonctionnement. Nous ne nous attarderons pas sur ces données puisqu’elles sont utilisées lors d’une phase d’introduction relativement longue et n’interviennent pas au cours de la session d’explication.

TAB. 6.1: Liste des sujets disponibles ainsi que les thèmes qui leur sont associés.

Sujet	Thèmes
REQUIN-CHASSE-CONCLUSION	REQUIN, CHASSE, CONCLUSION
REQUIN-ODORAT	REQUIN, ODORAT
REQUIN-TOUCHER	REQUIN, TOUCHER
REQUIN-VISION	REQUIN, VISION
REQUIN-REPRODUCTION	REQUIN, REPRODUCTION
REQUIN-ALIMENTATION	REQUIN, ALIMENTATION
REQUIN-DENTITION	REQUIN, DENTITION
REQUIN-RESPIRATION	REQUIN, RESPIRATION
REQUIN-SQUELETTE	REQUIN, SQUELETTE
REQUIN-FOND	REQUIN, FOND
REQUIN-PLEINEEAU	REQUIN, PLEINEEAU
REQUIN-INVENTAIRE	REQUIN, INVENTAIRE
PLATAX-AQUARIOPHILIE	PLATAX, AQUARIOPHILIE
PLATAX-CRITERES	PLATAX, CRITERES
PLATAX-ALIMENTATION	PLATAX, ALIMENTATION
PLATAX-HABITAT	PLATAX, HABITAT
PLATAX-ANATOMIE	PLATAX, ANATOMIE
PLATAX-NAGE	PLATAX, NAGE
NASO-REPRODUCTION	NASO, REPRODUCTION
NASO-GOUT	NASO GOUT
NASO-HABITAT-CRITERES	NASO, HABITAT, CRITERES
NASO-NAGE	NASO, NAGE

Suite à la page suivante...

Suite de la page précédente...	
Sujet	Thèmes
NASO-ALIMENTATION	NASO, ALIMENTATION
NASO-ANATOMIE-SCALPEL	NASO, ANATOMIE, SCALPEL
NASO-ANATOMIE-CORNE	NASO, ANATOMIE, CORNE
AQUARIUM	AQUARIUM

La liste des sujets et des thèmes définis est donnée par le tableau 6.1. Les éléments de discours (224 au total) associés à ces sujets comportent pour la plupart un script d’animation décrivant la lecture d’un fichier son généré automatiquement par synthèse vocale². Chaque élément de discours consiste en une ou plusieurs phrases d’une durée moyenne de 6 secondes pour un temps total d’environ 1200 secondes. A partir de ces éléments de discours 26 explications (une par sujet) ont été mis en place offrant différentes alternatives pour aborder le même thème. Le choix du chemin dans ces explications est effectué de manière aléatoire mais il est prévu, une fois le contenu stabilisé, d’y intégrer des fonctions de choix spécifiques tenant compte par exemple du temps restant avant la fin de la visite ou des sujets déjà abordés. A ceci s’ajoutent 16 transitions différentes adaptées à des cas particuliers ou des situations générales et utilisées lors du passage d’un sujet à un autre.

Le rendu est effectué par un moteur réalisé par l’entreprise Polymorph Software contrôlé par le module pédagogique à l’aide d’un protocole simple orienté sur la présentation d’information. Ce moteur est capable de gérer les déplacements de la représentation du guide virtuel dans l’environnement grâce à l’utilisation d’algorithmes de *steering* et de jouer différents types de documents multimédia (cf. annexe B).

6.2.4 Perception visuelle

L’environnement aquatique introduit de nombreux artefacts visuels difficiles à appréhender pour un système de vision artificielle. C’est le cas notamment des caustiques qui apparaissent et disparaissent en permanence sur l’ensemble du champ visuel. C’est également le cas pour les couleurs qui sont affectées par la masse d’eau que les rayons lumineux doivent traverser. Enfin le positionnement des caméras fait qu’il est possible de percevoir les visiteurs situés devant l’aquarium, ceux-ci pouvant alors cacher une partie du champ visuel.

Les différents comportements de perception n’ont pas les mêmes contraintes en terme de vitesse de traitement. Nous tirons ici parti des possibilités de parallélisation de notre architecture logicielle (cf. annexe C) en répartissant les comportements au sein de différents processus de la façon suivante :

- ▷ un processus se charge des comportements de vigilance et de suivi ainsi que des différents traitements associés ;
- ▷ un processus se charge du comportement de reconnaissance et des routines de perception associées.

² La synthèse vocale est utilisée afin de mettre au point le discours du guide. Dans la version finale (et notamment dans le cadre de la Nuit des Chercheurs) les sons joués ont été enregistrés par un être humain.

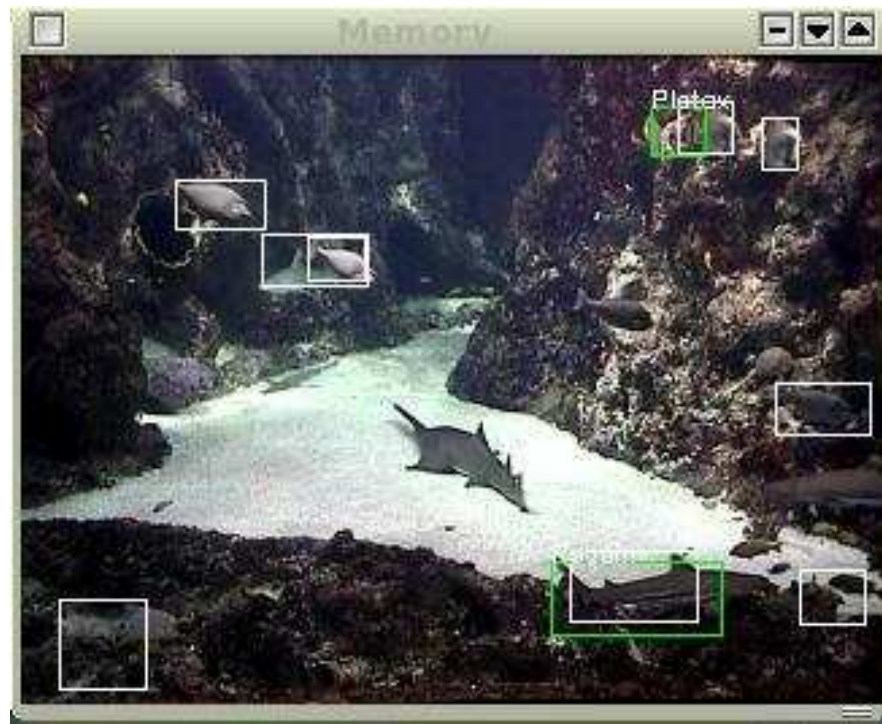


FIG. 6.9 – Capture d'écran du contenu de la représentation de l'environnement dans l'application au bassin d'Océanopolis.

La figure 6.9 montre l'état de la représentation de l'environnement créée par ce système de perception sur des images réelles.

La vigilance et le suivi s'effectuent sur des critères de mouvement, les poissons considérés pour les explications de la visite guidée se déplaçant en permanence dans l'environnement. La détection du mouvement se base sur l'utilisation conjointe d'un algorithme de soustraction de fond [KaewTraKulPong 01] et de différents filtres et opération morphologiques (érosion et dilation principalement) permettant d'atténuer les effets des caustiques. Il est envisagé d'ajouter d'autre types de détecteurs dans le futur mais les images fortement texturées du bassin requièrent malgré tout des traitements complexes. La détection s'effectue sur la totalité de l'image avec une fréquence d'application comprise entre 100 millisecondes et 10 secondes.

Les procédures, routines et stratégies nécessaires à l'identification des poissons ont été mises en œuvre à partir d'informations fournies par les biologistes d'Océanopolis. Ils nous ont décrits les caractéristiques des différentes espèces intéressantes de l'aquarium. Nous avons retenu les espèces suivantes dans le cadre de SIRENE :

▷ Platax :

- ◇ forme relativement circulaire néanmoins plus hauts que longs ;
- ◇ bandes noires verticales ;
- ◇ nagent en groupe ;



FIG. 6.10 – Les trois espèces considérées dans le cadre du projet SIRENE.

- ◇ se trouvent dans la zone centrale du champ visuel 90% du temps ;
- ◇ naviguent généralement entre le fond du bassin et environ 2 mètres sous la surface ;
- ◇ environ 60 cm de diamètre.

▷ Nasos :

- ◇ forme de ballon de rugby ;
- ◇ une double paire de scalpels bleus au niveau de la queue ;
- ◇ un corne plus ou moins prononcée à l'avant ;
- ◇ environ 40 cm de long.

▷ Requins :

- ◇ les plus imposants du bassin (environ 2,20 m pour le requin taureau) ;
- ◇ à ceci s'ajoutent des informations spécifiques sur des espèces particulières de requins (forme particulière par exemple).

Partant de ces informations, nous avons tenté de définir des propriétés et des classifieurs utilisables pour leur reconnaissance. Les différents processus de classification utilisés actuellement sont présentés dans la figure 6.11.

Les classifieurs utilisés reposent soit sur l'utilisation d'un seuil défini de façon empirique (zone d'évolution ou ratio de la boîte englobante par exemple) soit sur l'utilisation d'un algorithme de type SVM [Burges 98]. L'apprentissage et le paramétrage des différents classifieurs s'effectuent sur des images réelles extraites de façon automatique par le système de détection et de suivi. Le processus d'apprentissage ainsi que les différents classifieurs existants sont présentés dans l'annexe D.

Le tableau 6.2 offre un aperçu des résultats obtenus par le système de reconnaissance sur des enregistrements de l'aquarium. Ces taux correspondent à des évaluations du système effectuées en conditions réelles. Pour chaque espèce considérée nous présentons deux informations :

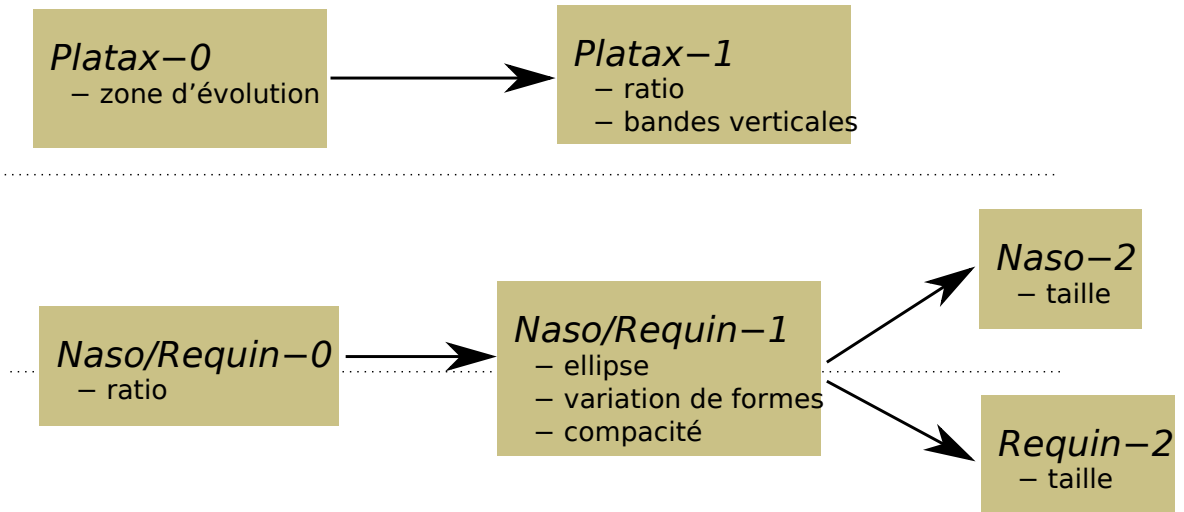


FIG. 6.11 – Classifieurs et propriétés pour la reconnaissance des espèces Platax, Naso et Requin.

Type	Précision	Rappel
Platax	0.714	0.476
Naso	0.697	0.436
Requin	0.693	0.436

TAB. 6.2 – Résultats du système de reconnaissance pour les trois espèces de poissons considérées dans le cadre du projet SIRENE.

- ▷ le taux de précision qui correspond au nombre d'entités correctement identifiées comme appartenant à une certaine classe par rapport au nombre total d'entités identifiées comme appartenant à cette classe, soit : $\frac{\text{vraipositifs}}{\text{vraipositifs} + \text{fauxpositifs}}$
- ▷ le taux de rappel qui correspond au nombre d'entités correctement identifiées comme appartenant à une certaine classe par rapport au nombre d'entités appartenant réellement à cette classe, soit : $\frac{\text{vraipositifs}}{\text{vraipositifs} + \text{fauxnegatifs}}$

Il apparaît lors de ces évaluations que le système de reconnaissance présente une précision acceptable alors que le taux de rappel est relativement faible (moins de 50%). Il est possible de faire varier ces résultats, l'augmentation du taux de rappel allant de paire avec une diminution du taux de précision. Les résultats présentés correspondent aux paramètres utilisés dans l'application réelle : un compromis entre le nombre d'entités identifiées et le nombre d'entités identifiées correctement. Dans le cadre du projet SIRENE, le taux de précision est essentiel : on doit être « certain » du poisson que l'on explique aux visiteurs.

Lors des expérimentations pour d'obtenir ces résultats, nous avons pu observer quelques problèmes :

- ▷ la nombre de cibles identifiées (c'est à dire allant jusqu'au terme d'un processus de classification) par rapport au nombre total de cibles suivies reste relativement faible. Seulement environ $\frac{1}{6}$ des cibles arrivent au terme de leur classification. Ce chiffre est à relativiser par rapport au fait que nous ne sommes capables d'identifier qu'une partie des espèces présentes dans l'aquarium ;
- ▷ les processus de classification que nous avons définis s'avèrent inefficaces dans certaines conditions particulières. Nous avons par exemple pu remarquer des erreurs de classification plus fréquentes sur les entités visuelles situées devant certaines parties de l'aquarium. L'utilisation de processus spécialisés pour certaines zones du champ visuel pourrait permettre de palier à cette difficulté, cette solution n'a néanmoins été que peu abordée ;
- ▷ le système de reconnaissance reste soumis à un fonctionnement correct du système de suivi : si une cible est perdue par le comportement de suivi, elle l'est aussi pour la reconnaissance.

6.2.5 Bilan

Dans cette section nous avons présenté l'application du guide virtuel autonome à la visite guidée d'un aquarium d'Océanopolis dans le cadre du projet ANR SIRENE. Nous avons tout d'abord précisé en quoi le guide virtuel que nous proposons s'avère intéressant pour répondre à un certain nombre de difficultés rencontrées par Océanopolis. Nous nous sommes ensuite attachés à présenter les deux aspects de cette application, à savoir la visite guidée et la perception visuelle.

L'aspect visite guidée consiste en l'ensemble des comportements et données de discours définies en coopération avec les responsables pédagogiques d'Océanopolis.

En ce qui concerne le système de perception visuelle, nous avons présenté les différentes espèces auxquelles nous nous sommes intéressés dans le cadre du projet SIRENE. Nous avons présenté des résultats sur les performances du système de reconnaissance et mis en avant les difficultés rencontrées sur ce point particulier de notre application.

Les sections suivantes présentent des expérimentation menées sur l'application dans des conditions contrôlées afin de mettre en évidence le comportement du système.

6.3 Expérimentations : la visite guidée

6.3.1 Introduction

Cette section présente des expérimentation relatives au processus de prise de décision du guide virtuel. Deux points sont abordés :

- ▷ la gestion des interruptions ;
- ▷ la sélection de sujet.

Ces expérimentations concernent la phase d'explication du comportement du guide virtuel qui est présenté dans la section 6.2.3. Les comportements locaux et les données de discours utilisés sont les mêmes que dans le cadre du projet SIRENE. Pour permettre un contrôle sur le déroulement de ces expérimentations et une reproductibilité des résultats, le système de perception visuelle est ici remplacé par des fichiers décrivant l'évolution de l'environnement au cours du temps. Cette évolution est précisée pour chacune des expérimentations présentées.

6.3.2 Gestion des interruptions

Le cœur de la visite guidée est la session d'explication. Au cours de cette étape, les comportements qui sont sélectionnés et exécutés permettent de présenter des informations au visiteur. Le guide virtuel alterne entre des phases de présentation d'informations et des phase de sélection des informations à présenter. La phase de présentation d'informations consiste en l'exécution d'un comportement permettant de dérouler un sujet d'explication. Une explication continue tant qu'elle n'est pas terminée ou qu'aucun événement susceptible de l'interrompre n'intervient. La phase de sélection des informations consiste en la réévaluation du sujet courant des explications. Cette réévaluation intervient lorsqu'un des événements suivants se produit :

- ▷ l'apparition d'une nouvelle entité visuelle ;
- ▷ la disparition de l'entité visuelle en cours d'explication ;
- ▷ la fin de l'explication en cours.

L'expérimentation que nous présentons ici décrit le déroulement d'une session d'explication d'une durée de 100 secondes. Le tableau 6.3 présente les différents événements qui se sont produits durant cette session ainsi que les dates (en secondes par rapport à la date de début de l'expérimentation) auxquelles ils se sont produits. Ce tableau présente parallèlement les différents éléments de discours présentés par le guide ainsi que la durée de chacun de ces éléments.

Le contenu de la représentation de l'environnement évolue de la façon suivante au cours de l'expérimentation :

- ▷ l'environnement ne contient aucun objet connu ;
- ▷ une entité de type « Platax » fait son apparition ;
- ▷ elle disparaît brièvement puis réapparaît ;
- ▷ une entité de type « Requin » vient s'ajouter ;
- ▷ le requin et le platax disparaissent au profit d'une entité de type « Naso ».

Le choix de l'explication courante est influencé par les « experts » suivants :

- ▷ un expert définissant un ordre de préférence *a priori* (*Requin* > *Naso* > *Platax*) ;
- ▷ un expert définissant une préférence pour les explications concernant un poisson présent dans la représentation de l'environnement.

Il est important de noter ici que nous nous intéressons plus à la façon dont les explications peuvent être interrompues qu'au sujet qui sont sélectionnés, ce point étant abordé dans l'expérimentation suivante (cf. section 6.3.3).

TAB. 6.3: Déroulement d'une visite sur une évolution de l'environnement déterminée.

Index	Date	Évènement	Discours
0	0	Début de session	-
1	10 ⁻⁴	Introduction	Salut ! C'est moi Didon, le diodon. Je suis un poisson porc-épic et j'habite ce lagon de l'océan Pacifique. (7.21)
2	7.21	-	Ici c'est mon domaine et je connais tout les poissons. (3.12)
3	10.33	-	Tu vois, je suis entouré de requins, des petits et des gros avec de longues dents pointues...mais, qu'ils se méfient, moi aussi j'ai des piquants pointus. (8.65)
4	13.40	Apparition Platax	-
5	19.00	Changement de sujet	Parlons des Platax. (2.22)

Suite à la page suivante...

Suite de la page précédente...			
Index	Date (s)	Évènement	Discours
6	21.22	-	Tu sais les amateurs d'aquariums adorent ces poissons à la nage majestueuse. Mais le problème c'est que leur croissance est tellement rapide que... eh bien il ont vite fait de remplir l'aquarium (10.04)
7	31.27	Fin explication	Les juvéniles ont une forme et une coloration très différentes de celles des adultes. (4.84)
8	31.40	Disparition Platax	-
9	35.90	Apparition Platax	-
10	36.13	Début explication	Les jeunes platax qui vivent dans les mangroves, entre les racines des arbres sont marrons ou noirs ; ça dépend des espèces. (7.00)
11	43.13	-	Avec leurs longues nageoires, on dirait des feuilles mortes tombées des arbres... (3.94)
12	44.40	Apparition Requin	-
13	47.09	Changement de sujet	Parlons des requins.
14	49.09	Fin explication	Le squelette des requins et des raies est fait de cartilage. (3.29)
15	52.40	Fin explication	En arrière de la tête les requins possèdent des ouvertures, cinq paires en général. (4.97)
16	57.39	Fin explication	Les requins sont armés...jusqu'aux dents : leur bouche est équipée de plusieurs rangées de dizaines de dents! (5.35)
17	62.76	Fin explication	De temps en temps, un petit poisson minuscule, le labre nettoyeur, vient leur faire les dents et la peau. (5.78477)
18	62.90	Disparition Platax ; Requin	-
19	67.40	Apparition Naso	-
20	68.56	Début explication	Ces drôles de poissons avec une bosse sur le front sont des nasos. Certains ont même une corne et ressemblent à des licornes. (6.81)
21	75.37	-	Ne me demande pas à quoi sert cette corne, je n'en sais rien. Mais je ne les ai jamais vu s'en servir pour se battre. (6.43)
22	81.82	Début explication	Les nasos ont une particularité anatomique, très utile. (3.23)
23	83.79	-	Ils possèdent 2 paires d'épines en avant de la queue pour se défendre. (3.79)
24	87.59	-	Ces épines osseuses sont des lames tranchantes triangulaires toujours pointées vers l'avant. (4.68)
Suite à la page suivante...			

Suite de la page précédente...			
Index	Date (s)	Évènement	Discours
25	92.28	-	Avec d'énergiques mouvements de la queue, ils peuvent infliger des blessures aux poissons et même aux pêcheurs. Et dire qu'il appartiennent au groupe des poissons chirurgiens ! Drôle de chirurgiens ces poissons. (10.58)
26	94.41	Disparition Naso	-
27	100.23	Fin de session	-
28	102.87	Conclusion	Au revoir... Si tu veux il y a plein d'autres animaux à découvrir dans les autres bassins du pavillon tropical. (5.77)

A l'indice 0, l'évènement de début de session déclenche le démarrage de la visite. Entre les indices 1 et 3 le guide est dans la phase d'introduction et se présente ainsi que son environnement d'une façon générale.

L'apparition d'un poisson de type « Platax » à l'instant 4 entraîne un changement de sujet à l'instant 5. Ce changement de sujet intervient environ 6 secondes après l'apparition du poisson dans l'environnement. Ce délai s'explique par la durée de l'élément d'explication en cours d'exécution (8.65 secondes), l'exécution d'un tel élément ne pouvant être interrompu.

Au indices 8 et 9 le poisson en cours d'explication disparaît brièvement. Cet évènement, bien qu'il ait été détecté, n'entraîne pas un changement de sujet ou une modification importante du discours car il intervient pendant l'exécution d'un élément de discours : le poisson est de nouveau présent dans l'environnement quand l'interruption devient possible.

Aux indices 11, 12 et 13 un phénomène similaire à celui s'étant produit au indices 4 et 5 intervient : un requin fait son apparition, celui-ci n'est pris en compte que 5 secondes plus tard du fait de l'élément d'explication en cours d'exécution. Ici nous pouvons également noter l'influence des différents experts sur le choix de l'explication. Deux entités (Requin et Platax) sont présentes dans la représentation de l'environnement, le guide choisi de parler du Requin car celui-ci est « préféré » par l'expert définissant un ordre *a priori*.

A l'indice 27, l'évènement de fin de session est déclenché, celui-ci n'est pris en compte qu'à l'indice 28 du fait de la durée de l'élément d'explication en cours d'exécution.

6.3.3 Sélection des explications

Une session d'explication est interrompue en permanence lorsque des entités visuelles apparaissent ou disparaissent de la représentation de l'environnement ou que l'explication en cours arrive à son terme. Lorsque l'une de ces conditions intervient, le sujet courant traité par le guide virtuel est réévalué et un nouveau sujet est éventuellement sélectionné. Le processus de sélection est sous la responsabilité de l'ensemble des experts actifs à cet instant donné. Cet ensemble d'experts influence le choix du sujet des explications en fonction de différents points de vue.

Chaque expert est une fonction qui se voit passer en paramètre l'ensemble des sujets disponibles (c'est à dire les sujets non terminés à un instant donné). Pour chaque élément de cet ensemble, un expert a le choix entre trois actions :

- ▷ attribuer une voix à l'élément en question ;
- ▷ faire usage d'un droit de veto sur l'élément en question ;
- ▷ ne rien faire, c'est à dire laisser inchangé le nombre de voix de l'élément en question.

Nous avons mené des expérimentations quant au choix des experts actifs et à leur influence sur le déroulement de la visite au sens de la sélection de sujet. Pour cela nous avons défini les experts suivants :

- ▷ **entity-expert** dont le rôle est de favoriser les sujets pour lesquels il existe une entité visuelle pertinente dans la représentation courante de l'environnement ;
- ▷ **not-started-expert** dont le rôle est de favoriser les sujets qui n'ont pas encore été abordés ;
- ▷ **interrupted-expert** dont le rôle est de favoriser les sujets qui ont été suspendus ;
- ▷ **no-entity-expert** dont le rôle est de favoriser les sujets ne nécessitant aucune entité visuelle spécifique lorsque la représentation de l'environnement ne contient aucune entité connue ;
- ▷ **minimum-time-expert** dont le rôle est de favoriser le sujet qui vient d'être interrompu si la durée de son exécution jusqu'à présent est inférieure à une durée déterminée ;
- ▷ **same-topic-expert** dont le rôle est de favoriser les sujet ayant des thèmes communs avec le sujet courant ;
- ▷ **different-topic-expert** dont le rôle est de favoriser les sujets ne possédant aucun thème commun avec le sujet courant ;
- ▷ **order-expert** dont le rôle est de définir un ordre *a priori* sur les thèmes devant être abordés.

Les différents experts que nous avons mis en place lors de cette expérimentation ne sont bien entendu pas les seuls pouvant exister, néanmoins il nous a été possible de mettre en évidence plusieurs *catégories* parmi les experts pouvant être définis :

- ▷ les experts en rapport avec le contenu de la représentation de l'environnement. Dans le cadre de notre expérimentation ce sont les experts suivants : **entity-expert**, **no-entity-expert** ;
- ▷ les experts en rapport avec le déroulement de la visite. Ce sont ici les experts suivants : **not-started-expert**, **interrupted-expert**, **same-topic-expert** et **different-topic-expert** ;

- ▷ les experts en rapport avec l'aspect temporel de la visite. Ici cela concerne l'expert `minimum-time-expert` ;
- ▷ les experts permettant de représenter des connaissances *a priori* tels que ici : `order-expert`. Ces experts ne tiennent pas compte du contexte courant de la visite.

Dans le cadre de notre expérimentation, nous avons voulu mettre en évidence les différentes actions possibles pour le vote des experts en définissant une version spéciale pour deux des experts présentés précédemment : `entity-expert` et `minimum-time-expert`. A partir de ces deux experts, nous avons définis les experts suivants : `entity-veto-expert` et `minimum-time-veto-expert` qui reprennent le fonctionnement des experts « classiques » à ceci prêt qu'il font usage d'un droit de veto pour les sujet qu'il ne souhaitent pas mettre en valeur. Ainsi `entity-veto-expert` applique son veto pour tous les sujets qui ne sont pas en rapport avec des entités connues présentes dans la représentation de l'environnement lorsque celle-ci contient au moins une entité connue. `minimum-time-veto-expert` quant à lui applique son veto pour le sujet courant si le temps minimum est écoulé obligeant ainsi le guide à sélectionner un sujet différent.

A partir de ces différents experts, nous avons mis en place quatre configurations différentes permettant de mettre en valeur l'influence de tel ou tel expert sur la structure globale de la visite guidée :

- ▷ une configuration dite *idéale* dans laquelle le déroulement de la visite prend en compte à la fois les aspects connaissances *a priori*, le contenu de la représentation de l'environnement et le contexte de la visite. Cette configuration fait appel aux experts suivants :
 - ◇ `entity-expert` ;
 - ◇ `no-entity-expert` ;
 - ◇ `same-topic-expert` ;
 - ◇ `order-expert` ;
 - ◇ `minimum-time-expert` ;
- ▷ une configuration dite *veto* qui est similaire à la configuration *idéale* à ceci prêt que les experts qui le peuvent sont remplacés par leur version faisant appel à leur droit de veto. Ainsi `entity-expert` et `minimum-time-expert` sont remplacés par `entity-veto-expert` et `minimum-time-veto-expert` ;
- ▷ une configuration dite *volatile* dans laquelle le guide virtuel a tendance à changer de sujet lors de chaque interruption. Cette configuration repose sur l'utilisation des experts suivants :
 - ◇ `entity-expert` ;
 - ◇ `no-entity-expert` ;
 - ◇ `different-topic-expert` ;
 - ◇ `order-expert` ;

- ▷ une configuration dite *réactive* dans laquelle le guide virtuel réagit immédiatement aux modifications de la représentation de l'environnement. Cette configuration est basée sur la collaboration des experts suivants :

- ◊ `entity-expert` ;
- ◊ `no-entity-expert` ;
- ◊ `order-expert`.

Dans chacune de ces configuration, `order-expert` définit l'ordre suivant sur une partie des thèmes disponibles : *Requin* > *Naso* > *Platax*.

Pour cette expérimentation l'évolution du contenu de la représentation de l'environnement à été créée dans un fichier sur une durée d'environ 450 secondes. Cette évolution est présentée dans le tableau 6.4.

TAB. 6.4: Évolution de la représentation de l'environnement.

Date (s)	Contenu de la représentation
0	Aucune entité connue
35	Platax
55	Aucune entité connue
60	Platax
70	Aucune entité connue
75	Platax
95	Aucune entité connue
125	Platax, Requin
185	Requin
215	Aucune entité connue
245	Naso
260	Aucune entité connue
265	Naso
275	Platax, Naso
325	Platax, Naso, Requin
375	Platax
395	Aucune entité connue
400	Platax
405	Naso
410	Aucune entité connue
415	Naso
435	Aucune entité connue

A partir de cette évolution de l'environnement, nous avons exécuté une session d'explication pour chacune des configurations existante. Le tableau 6.5 présente le déroulement de la visite pour chacune des configurations au cours de cette exécution. Dans ce tableau, les traits horizontaux épais représentent les instants auxquels un changement se produit dans la représentation de l'environnement. Le mot clé **END** permet de représenter la fin d'une ex-

plication complète. Chaque colonne correspond à une configuration particulière et présente l'enchaînement des différents sujets au cours du temps.

Il est important de noter que lors de ces expérimentations nous avons volontairement annulé tout choix aléatoire afin de pouvoir comparer correctement les effets des différentes configurations d'experts. Ainsi lorsque plusieurs sujets possèdent un nombre de voix équivalent c'est le premier (suivant l'ordre du tableau 6.1) qui est sélectionné. De même au sein des scénarios, c'est toujours la première option qui est sélectionnée lorsque plusieurs sont possibles, assurant ainsi un déroulement (et donc une durée totale d'exécution) identique pour toutes les explications pour chaque exécution.

TAB. 6.5: Déroulement de la visite pour différentes configurations d'experts.

Date	ideal	réactive	veto	volatile
22	AQUARIUM	AQUARIUM	AQUARIUM	AQUARIUM
44	PLATAX-AQUARIOPHILIE	PLATAX-AQUARIOPHILIE	-	PLATAX-AQUARIOPHILIE
49	-	-	PLATAX-AQUARIOPHILIE	-
58	-	AQUARIUM	-	-
59	-	-	-	AQUARIUM
62	END	-	PLATAX-CRITERES	-
64	PLATAX-CRITERES	-	-	-
70	-	PLATAX-AQUARIOPHILIE	-	-
71	-	-	-	PLATAX-AQUARIOPHILIE
76	-	-	END	-
77	-	END	-	-
78	END	-	PLATAX-AQUARIOPHILIE	END
79	-	PLATAX-CRITERES	-	-
80	PLATAX-ALIMENTATION	-	-	PLATAX-CRITERES
85	-	-	END	-
87	-	-	PLATAX-ALIMENTATION	-
91	END	-	-	-
93	PLATAX-HABITAT	END	-	-
94	-	-	-	END
95	-	PLATAX-ALIMENTATION	-	-
96	-	-	END	PLATAX-ALIMENTATION
97	-	-	REQUIN-ALIMENTATION	-
105	-	END	-	END
107	-	-	-	AQUARIUM
110	-	AQUARIUM	-	REQUIN-CHASSE-CONCLUSION

Suite à la page suivante...

Suite de la page précédente...

Date	ideal	réactive	veto	volatile
113	-	-	REQUIN-CHASSE-CONCLUSION	-
117	-	END	-	-
118	-	REQUIN-CHASSE-CONCLUSION	-	-
119	-	-	-	AQUARIUM
124	-	-	REQUIN-ODORAT	END
125	-	-	-	REQUIN-CHASSE-CONCLUSION
126	REQUIN-CHASSE-CONCLUSION	-	-	-
133	-	END	-	-
134	-	REQUIN-ODORAT	-	-
136	-	-	-	END
138	-	-	-	REQUIN-ODORAT
141	END	-	-	-
143	REQUIN-ODORAT	-	END	-
145	-	-	REQUIN-CHASSE-CONCLUSION	-
149	-	-	END	-
151	-	END	REQUIN-TOUCHER	-
153	-	REQUIN-TOUCHER	-	-
157	-	-	-	END
159	END	-	-	REQUIN-TOUCHER
161	REQUIN-TOUCHER	-	-	-
172	-	-	END	-
174	-	-	REQUIN-ELECTRORECEPTION	-
175	-	END	-	-
176	-	REQUIN-ELECTRORECEPTION	-	-
180	-	-	-	END
182	-	-	-	REQUIN-ELECTRORECEPTION
183	END	-	-	-
184	REQUIN-ELECTRORECEPTION	-	-	-
189	-	-	REQUIN-VISION	-
210	-	END	-	-
211	-	REQUIN-VISION	-	-
215	-	-	-	END
218	END	-	-	-
219	REQUIN-VISION	-	-	-
222	-	-	-	PLATAX-HABITAT
227	-	-	REQUIN-ELECTRORECEPTION	REQUIN-VISION
235	-	-	REQUIN-VISION	-
238	-	-	-	PLATAX-HABITAT

Suite à la page suivante...

Suite de la page précédente...

Date	ideal	réactive	veto	volatile
239	-	-	REQUIN-ELECTRORECEPTION	-
244	-	-	REQUIN-VISION	-
246	-	-	END	REQUIN-VISION
248	-	NASO-REPRODUCTION	NASO-REPRODUCTION	-
254	-	-	-	NASO-REPRODUCTION
256	NASO-REPRODUCTION	-	-	-
263	-	REQUIN-VISION	REQUIN-REPRODUCTION	REQUIN-VISION
268	-	NASO-REPRODUCTION	NASO-REPRODUCTION	-
274	-	-	-	NASO-REPRODUCTION
276	-	-	END	-
278	-	END	NASO-GOUT	-
279	END	-	-	-
280	-	NASO-GOUT	-	-
281	NASO-GOUT	-	-	PLATAX-HABITAT
314	-	-	END	-
315	-	END	NASO-HABITAT-CRITERES	-
316	END	-	-	-
317	-	NASO-HABITAT-CRITERES	-	-
318	NASO-HABITAT-CRITERES	-	-	-
329	-	-	REQUIN-ELECTRORECEPTION	-
330	-	REQUIN-VISION	-	-
331	REQUIN-VISION	-	-	END
332	-	END	-	REQUIN-VISION
334	-	REQUIN-REPRODUCTION	-	-
336	END	-	-	-
338	REQUIN-REPRODUCTION	-	-	-
340	-	-	END	-
342	-	-	REQUIN-REPRODUCTION	-
354	-	-	-	END
356	-	-	-	REQUIN-REPRODUCTION
377	PLATAX-HABITAT	-	-	-
378	-	-	PLATAX-HABITAT	-
382	-	-	-	PLATAX-ANATOMIE
391	-	PLATAX-HABITAT	-	-
397	-	REQUIN-REPRODUCTION	PLATAX-ANATOMIE	-
398	-	-	-	REQUIN-REPRODUCTION
403	-	-	-	PLATAX-ANATOMIE

Suite à la page suivante...

Suite de la page précédente...

Date	ideal	réactive	veto	volatile
406	END	-	-	-
407	-	PLATAX-HABITAT	-	-
408	NASO-HABITAT-CRITERES	-	-	-
409	-	-	NASO-ANATOMIE-SCALPEL	-
410	-	-	-	NASO-REPRODUCTION
415	-	REQUIN-REPRODUCTION	-	-
416	-	-	-	REQUIN-ALIMENTATION
417	END	-	-	-
418	NASO-NAGE	-	-	-
423	-	NASO-HABITAT-CRITERES	-	NASO-REPRODUCTION
425	END	-	-	-
427	NASO-ALIMENTATION	-	-	END
429	-	-	-	NASO-GOUT
432	-	END	-	-
433	-	NASO-NAGE	-	-
438	-	-	-	REQUIN-REPRODUCTION
439	-	REQUIN-REPRODUCTION	-	-
440	REQUIN-ALIMENTATION	-	-	-
447	-	-	-	PLATAX-ANATOMIE
449	-	-	PLATAX-ANATOMIE	-

La configuration *idéale* présente une certaine réactivité par rapport au contenu de l'environnement. C'est le cas par exemple à l'instant 44 lorsque le sujet **PLATAX-AQUARIOPHILIE** est sélectionné. Ce choix intervient lors de l'apparition d'une entité visuelle de type **Platax** dans la représentation de l'environnement. C'est également le cas aux instants 126, 256, 377, 408 et 440. Parallèlement, cette configuration permet également une certaine continuité en favorisant les sujets possédant un thème commun et en favorisant une durée minimum de 10 secondes pour les sujets commencés. Cela a pour effet d'augmenter la cohérence globale de la session d'explication. On peut observer au cours de son déroulement une tendance à ignorer les changements « locaux » au sein de la représentation. On peut observer ce phénomène lors des disparitions brèves des entités visuelles de type **Platax** au premiers instants de la représentation de l'environnement comme lors des instants 44 et 64. L'utilisation de connaissances *a priori* (l'ordre des sujets en rapport avec les entités visuelles connues) permet également de favoriser les sujets en rapport avec le thème **Requin**. Ceci peut être observé par exemple à l'instant 126 lorsque des entités de types **Platax** et **Requin** sont présentes dans la représentation de l'environnement.

La configuration *veto* possède un comportement similaire à celui de la configuration *idéale*

à quelques différences prêt liées à la présence du droit de veto. L'expert associé au temps passé sur le sujet courant impose par exemple le changement de sujet à l'instant 62. Celui associé aux entités visuelles connues présentes dans l'environnement impose le retour sur le sujet NASO-REPRODUCTION à l'instant 268 après que le veto sur le temps ait imposé un passage sur le sujet REQUIN-VISION lors de la brève disparition de l'entité visuelle de type **Platax**.

La configuration *réactive* reprend la configuration idéale sans les aspects de continuité (*i.e.* uniquement les caractéristiques réactives). Ceci s'observe dès les premiers instants de la visite guidée lors de l'instant 58 par exemple et de la disparition brève de l'entité visuelle de type **Platax**. Cette situation est d'ailleurs reproduite aux instants 263 et 387. De plus la présence de **entity-modifier-expert** et de **order-expert** permet d'assurer un ordre et de favoriser les sujets en rapport avec le contenu de l'environnement.

La configuration dite *volatile* reprend les experts de la configuration *réactive* à ceci prêt que l'expert **different-topic-expert** amène une tendance à sélectionner des sujets en rapport avec des thèmes différents de ceux du sujet courant. Ainsi lorsque la session d'explications est interrompue le sujet qui est sélectionné pour continuer la visite guidée comporte obligatoirement un thème différent de ceux du sujet courant. Cela s'observe lors de chaque changement de sujet comme par exemple aux instants 59, 107 ou encore 119.

6.3.4 Bilan

Dans cette section nous avons présenté deux expérimentations concernant la description du réel par le guide virtuel.

Au cours de la première expérimentation nous avons mis en évidence le fonctionnement global d'une session d'explication et la façon dont la visite se déroule en fonction de l'environnement. Ceci nous a permis de montrer comment le guide réagit aux événements susceptibles de se produire dans l'environnement en tenant compte de l'atomicité des éléments de discours. Cela nous a également permis de mettre en évidence les ordres de grandeur que nous considérons comme étant le temps réel pour le discours (la durée des différents éléments de discours).

La seconde expérimentation nous a permis de mettre en évidence l'influence de différentes configurations d'experts sur le choix des explications fournies par le guide virtuel. Il ressort de ces expérimentations qu'il pourrait être intéressant d'étudier certains aspects de la sélection de sujet comme par exemple l'apparition de schémas dans le déroulement de la visite tels que l'alternance entre deux mêmes sujets dans certaines configurations. Le nombre et le type des experts utilisés mériteraient également d'être étendu ainsi que l'évolution de la représentation de l'environnement modifiée afin de mettre en évidence des situations problématiques. Le contrôle offert par le mécanisme des experts que nous venons de mettre en évidence nous semble cependant satisfaisant du point de vue de notre application et son approfondissement nécessiterait certainement l'ajout de nouvelles connaissances comme par exemple l'aspect pédagogique ou les intérêts du visiteur que nous avons volontairement mis de côté ici.

6.4 Expérimentations : perception visuelle

6.4.1 Introduction

Dans cette section nous présentons des expérimentations concernant le système de perception visuelle et les mécanismes de contrôle attentionnel que nous avons mis en place.

Afin de mettre en évidence les propriétés du système de perception, nous avons mené ces expérimentations en utilisant une vidéo contenant des objets mobiles artificiels facilement identifiables. Les objets présents dans cette vidéo possèdent des trajectoires linéaires ce qui facilite leur suivi. Nous tentons ainsi de nous détacher des problématiques de vision par ordinateur « classique »³ afin de nous concentrer sur les propriétés de notre système par rapport à une capacité de calcul limitée et de mettre en évidence l'intérêt du mécanisme de contrôle basé sur l'attention visuelle. Cette vidéo contient des stimuli visuels simples symbolisant 3 types d'objets différents que nous avons nommés **red**, **green** et **blue** possédant respectivement les couleurs rouge, verte et bleue (figure 6.12). Lorsqu'un objet sort du cadre de la vidéo il réapparaît du côté opposé (espace « torique »), ainsi à tout instant 10 entités sont visibles (5 entités de type **red**, 3 de type **green** et 2 de type **blue**).

6.4.2 Représentation de l'environnement

Lors de la première phase d'expérimentation nous avons voulu mettre en évidence le fonctionnement du système sans aucune influence extérieure, *i.e.* en l'absence d'effet du mécanisme attentionnel.

Les différents tests que nous avons menés sur cette vidéo impliquent les configurations suivantes des différents objets du système de perception visuelle :

- ▷ le comportement de **vigilance** qui à la charge de la détection des différentes entités de l'environnement se base sur trois routines de détection :
 - ◊ Détection des objets de couleur rouge. Basée sur l'utilisation d'une carte de propriété (**red-prop**). Les fréquences minimum et maximum d'exécution sont respectivement de 100 et 5000 millisecondes ;
 - ◊ Détection des objets de couleur verte. Basée sur l'utilisation d'une carte de propriété (**green-prop**). Les fréquences minimum et maximum d'exécution sont respectivement de 100 et 5000 millisecondes ;
 - ◊ Détection des objets de couleur bleue. Basée sur l'utilisation d'une carte de propriété (**blue-prop**). Les fréquences minimum et maximum d'exécution sont respectivement de 100 et 5000 millisecondes ;

³ Le problème de la reconnaissance dans le cadre d'un environnement réel reste complexe au delà des problématiques de contrôle que nous présentons ici.

- ▷ le comportement de **reconnaissance** fait appel à trois classifieurs, chacun ayant la charge d'identifier un type particulier d'entité visuelle (**red**, **green** et **blue**).

Cet ensemble d'objet correspond à la configuration minimale que nous avons identifiée pour la mise en place d'un système de perception capable de construire une représentation exploitable par le processus de prise de décision. La configuration utilisée dans l'application (cf. section 6.2) ne se différencie que par les routines utilisées et une distribution du système sur plusieurs processus : les comportements de perception sont identiques.

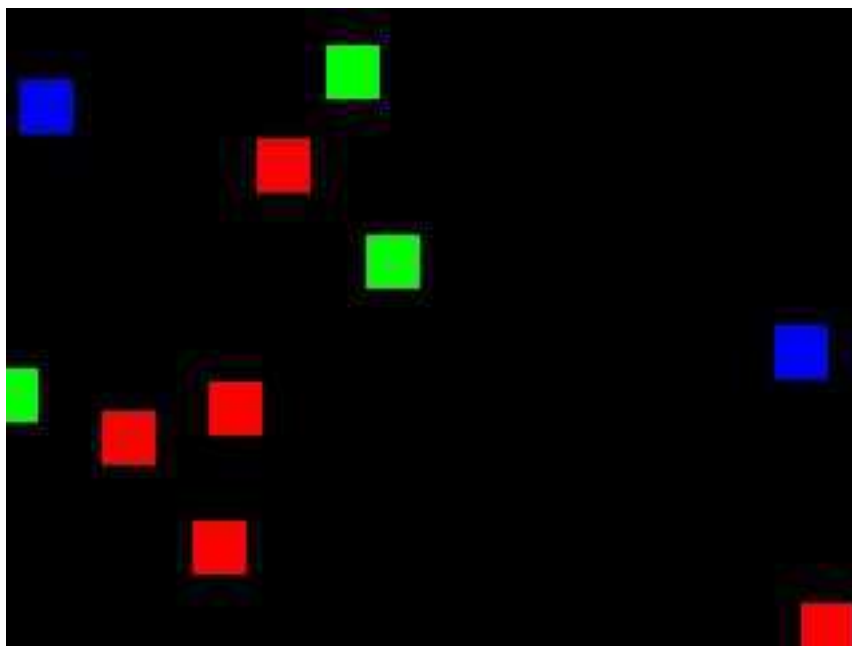


FIG. 6.12 – Capture d'écran de la vidéo utilisée pour les expérimentations sur le système de perception visuelle.

Partant de cette configuration nous avons observé différentes propriétés de notre système en absence de tout contrôle attentionnel, *i.e.* les valeurs d'intérêt sont équilibrées au sein des différents comportements de perception.

Lors de cette expérimentation, nous avons fait varier la contrainte de puissance de calcul disponible en modulant le nombre minimum d'images par seconde devant être respecté par le module de traitement d'images. Nous avons exécuté le système pour des limitations de 10, 15 et 20 images par seconde et ce pendant une durée de 100 secondes. Cette limitation fonctionne de la façon suivante : à chaque pas d'exécution du module de traitement d'images le temps depuis la dernière exécution est calculé afin de déterminer le temps disponible pour effectuer les calculs en considérant la limitation imposée. Puis après chaque traitement ce temps est réévalué. Si le temps restant est insuffisant, une erreur est renvoyée aux routines ayant demandé les traitements encore non effectués et ceux-ci sont annulés. Un tel mécanisme est utile afin de garantir le bon fonctionnement de certains comportements de perception exigeants en terme de mise à jour tels que le suivi et ainsi assurer une bonne cohérence entre l'environnement et sa représentation. Pour chacune des exécutions (10, 15 et 20 images par seconde) nous avons observé les paramètres suivants :

- ▷ nombre total d'entités détectées et suivies ;
- ▷ nombre total d'entités identifiées ;
- ▷ nombre d'entités pour chaque type existant (**red**, **green** et **blue**).

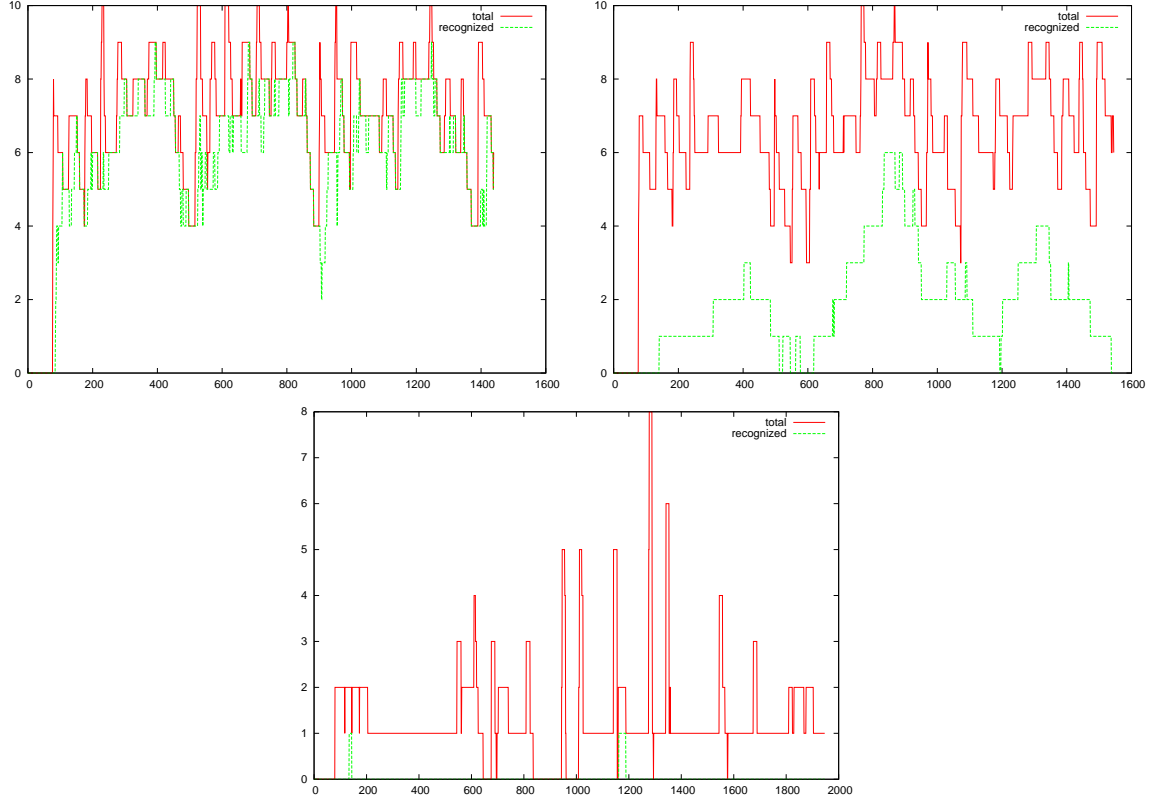


FIG. 6.13 – Nombre de cibles suivies et reconnues pour des limitations à 10, 15 et 20 images par secondes pour l'objet effectuant les traitements.

La figure 6.13 présente ainsi sur le même graphique les variations du nombre total d'entités suivies (et donc préalablement détectées) et le nombre total d'entités identifiées à un instant donné. Ici nous ne considérons pas les erreurs d'identification qui sont d'ailleurs inexistantes étant donné la simplicité des entités utilisées pour le test. Les trois graphiques présentés correspondent (de gauche à droite et de haut en bas) aux limitations à 10, 15 et 20 images par seconde.

Sur le premier graphique on peut observer un nombre d'entités détectées et suivies proche du nombre total d'entités présentes dans l'environnement, les légères variations s'expliquent par la disparition et la réapparition des différentes entités lorsqu'elle atteignent les limites de l'environnement (figure 6.14). On peut également observer des variations du nombre d'entités identifiées similaires aux variations du nombre d'entités suivies. Ces résultats s'expliquent par le fait que la limite à 10 images par secondes laisse au système de perception suffisamment de temps pour effectuer tous les traitements nécessaires au suivi et à l'identification de toutes les entités en mémoire.

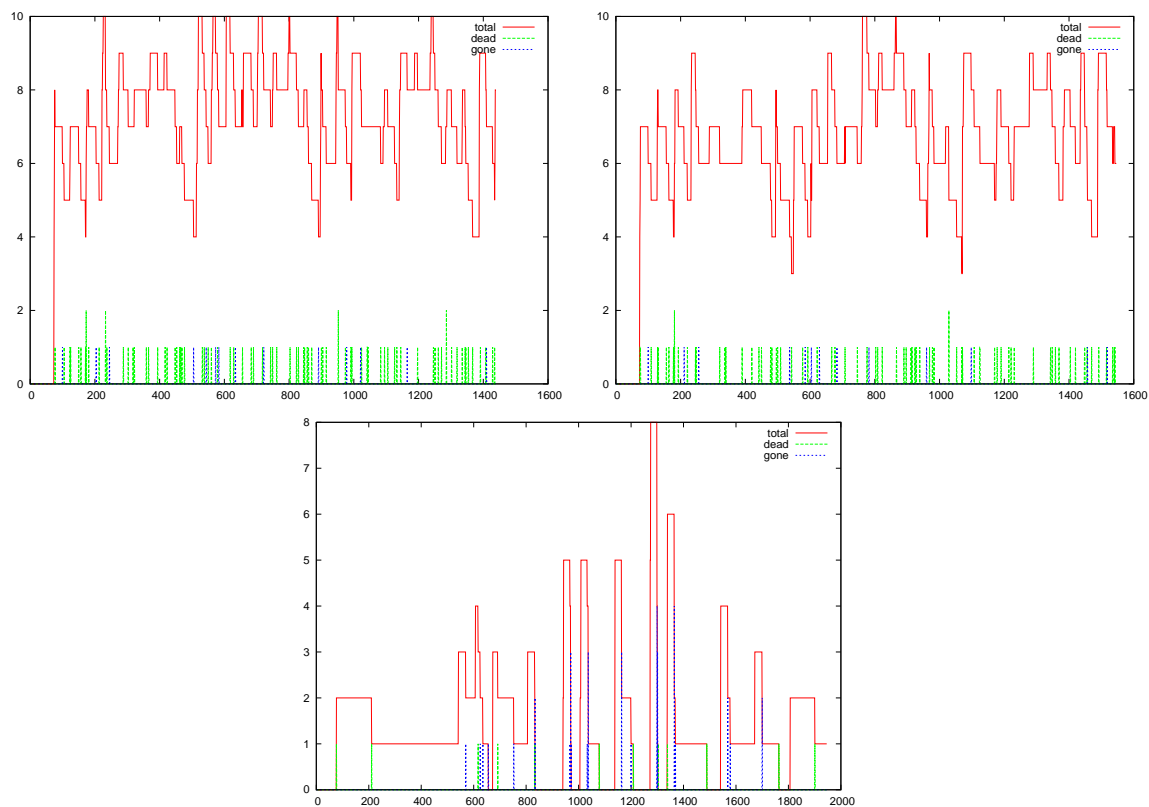


FIG. 6.14 – Nombre de cibles suivies et disparues/perdues au cours du temps pour des limitations de 10, 15 et 20 images par seconde.

Lorsque la limitation passe à 15 images par seconde (second graphique) on peut observer une légère variation au niveau du nombre d'entités suivie qui s'explique par le fait que le temps imparti pour effectuer les traitements n'est pas suffisant pour effectuer l'ensemble des traitements nécessaires à la construction d'une représentation complète de l'environnement. On observe également un nombre d'entités identifiées beaucoup plus faible que le nombre d'entités suivies. Ceci s'explique ici encore pas le temps insuffisant pour effectuer la totalité des traitements requis. Le maintien d'un nombre de cibles suivies relativement acceptable s'explique par le fonctionnement du comportement de suivi qui tente d'effectuer des mises à jour extrêmement fréquemment et possède un mécanisme de prédiction basé sur l'utilisation d'un filtre de Kalman qui permet de passer outre l'absence de mise à jour des propriétés géométriques et dynamique d'une entité pendant quelques instants.

Si la limitation passe à 20 images par secondes, le temps disponible pour effectuer les différents traitements d'image nécessaires devient insuffisant. Cela se répercute directement sur les variables observées. Ainsi le nombre d'entités suivies chute de façon dramatique avec quelques sursauts qui s'expliquent par les phases de détection lors desquelles de nombreuses entités sont ajoutées en mémoire, celles-ci disparaissant presque immédiatement du fait de l'absence de mise à jour durable. Le nombre d'entités identifiées quant à lui est proche de zéro, les traitements ne pouvant être effectués dans le temps imparti.

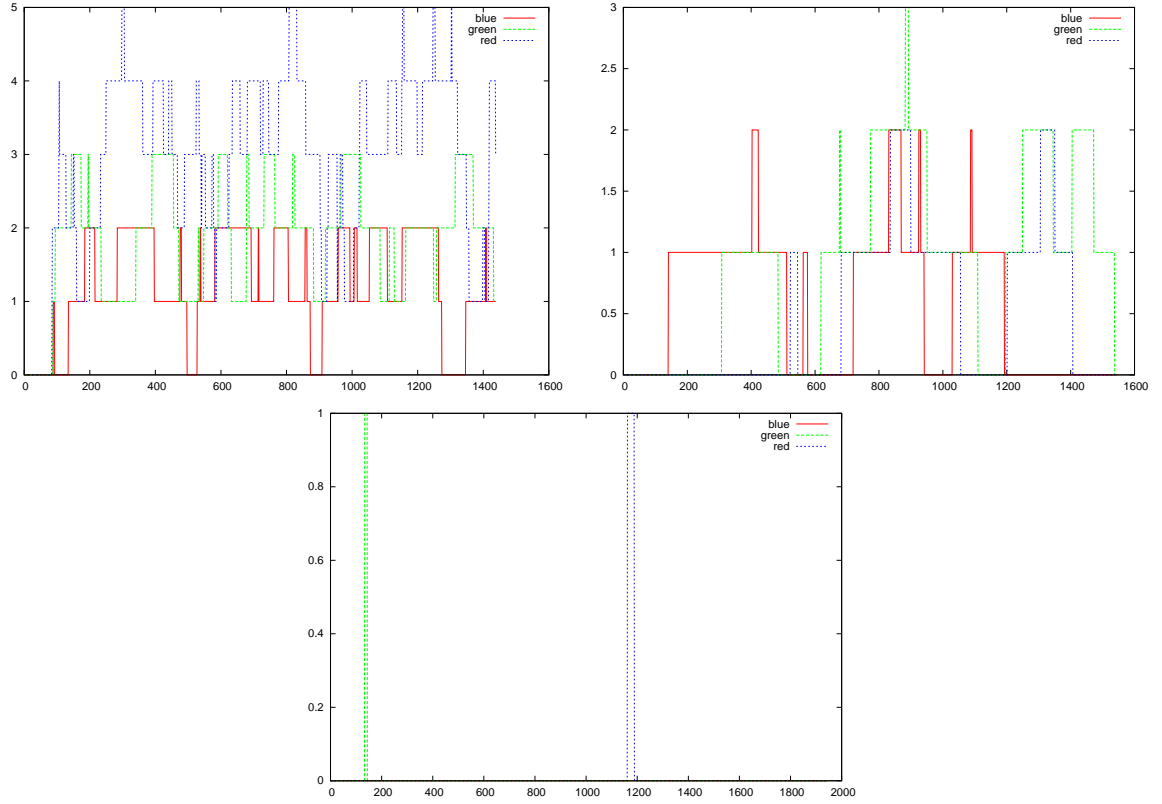


FIG. 6.15 – Nombre de cibles reconnues pour chacun des types possibles (red, green et blue) pour des limitations à 10, 15 et 20 images par secondes.

La figure 6.15 présente le nombre de cibles identifiées pour chacun des types présents dans l'environnement. Les trois graphiques correspondent ici aussi aux différentes limitations mises en place (10, 15 et 20 images par seconde). On observe ici les mêmes phénomènes que pour les paramètres précédents à savoir : une diminution du nombre d'entités identifiées lorsque la limite en terme de puissance de calcul augmente et des variations locales dues aux apparitions et disparitions fréquentes d'entités dans l'environnement. On peut également observer ici un certain équilibre entre les différents types d'entités identifiées : en dehors des différences dues à un nombre total différent pour chaque type d'entités le système ne possédant aucune information *a priori* sur le type des entités à identifier produit un résultat équivalent pour chacun des types existant.

6.4.3 Contrôle attentionnel

Lorsque les ressources de traitement d'image sont insuffisantes seule une partie de l'environnement peut être représentée de façon correcte. Le contenu de cette représentation dépend alors des traitements qui ont été déclenchés et ont pu être menés à leur terme. Dans de telles conditions, les informations présentes dans la représentation peuvent s'avérer peu pertinentes du point de vue du processus de prise de décision et de la description de l'environnement par

le guide virtuel. Le mécanisme de contrôle attentionnel permet d'influencer les traitements qui vont être effectués afin d'obtenir dans la représentation de l'environnement les informations que recherche le processus de prise de décision. Pour cela, des connaissances *a priori* sont utilisées afin de sélectionner des traitements qui permettent d'obtenir les informations souhaitées. Le mécanisme de contrôle attentionnel permet de modifier le processus de sélection des routines par les comportements par la prise en compte d'une valeur d'intérêt.

Afin de mettre en évidence les propriétés de ce mécanisme de contrôle, nous avons exécuté le système sur des sessions d'une durée de 200 secondes. Durant cette exécution, nous avons déclenché le mécanisme de contrôle en portant l'attention du système de perception sur différents types d'entités de la façon suivante :

- ▷ les entités de type **red** peu après le début de la session ;
- ▷ environ 30 secondes après avoir stoppé l'intérêt sur ce type d'entité nous avons porté l'attention du système sur des entités de type **green**, ce délai permettant au système « d'oublier » l'influence précédente grâce au mécanisme d'atténuation ;
- ▷ puis, sans laisser le mécanisme d'atténuation faire son œuvre, nous avons porté l'attention du système sur des entités de type **blue**.

Nous avons répété l'opération pour des limitations de 10 et 15 images par seconde. La configuration à 20 images par seconde a été écartée, les résultats obtenus donnant lieu à des remarques identiques à celle présentées pour la limitation à 15 images par seconde⁴ Les figures 6.16 (limitation à 10 images par secondes) et 6.17 (limitation à 15 images par seconde) présentent la mesure de différents paramètres du système au cours de ces expérimentations (de haut en bas et de gauche à droite) :

- ▷ le nombre total d'entités suivies et identifiées ;
- ▷ le nombre d'entités de chaque type ;
- ▷ la moyenne mobile du nombre d'exécution des classifieurs existants (un pour chaque type d'entité) ;
- ▷ les variations du temps entre deux détection d'un certain type d'entité ;
- ▷ l'évolution de l'intérêt au niveau du comportement de vigilance ;
- ▷ l'évolution de l'intérêt au niveau du comportement de reconnaissance (intérêt pour les types).

Lorsque la limitation des capacités de traitements est fixée à 10 images par seconde (figure 6.16), l'influence du mécanisme de contrôle attentionnel est relativement limitée du fait que les ressources disponibles sont suffisantes pour effectuer la totalité des traitements. Il

⁴ Il est néanmoins important de noter qu'avec cette configuration écartée les traitements restent trop coûteux pour permettre l'identification et le suivi efficace d'un nombre conséquent d'entités. Ceci provenant du fait que les traitements restent malgré tout trop coûteux pour être effectués de façon correcte.

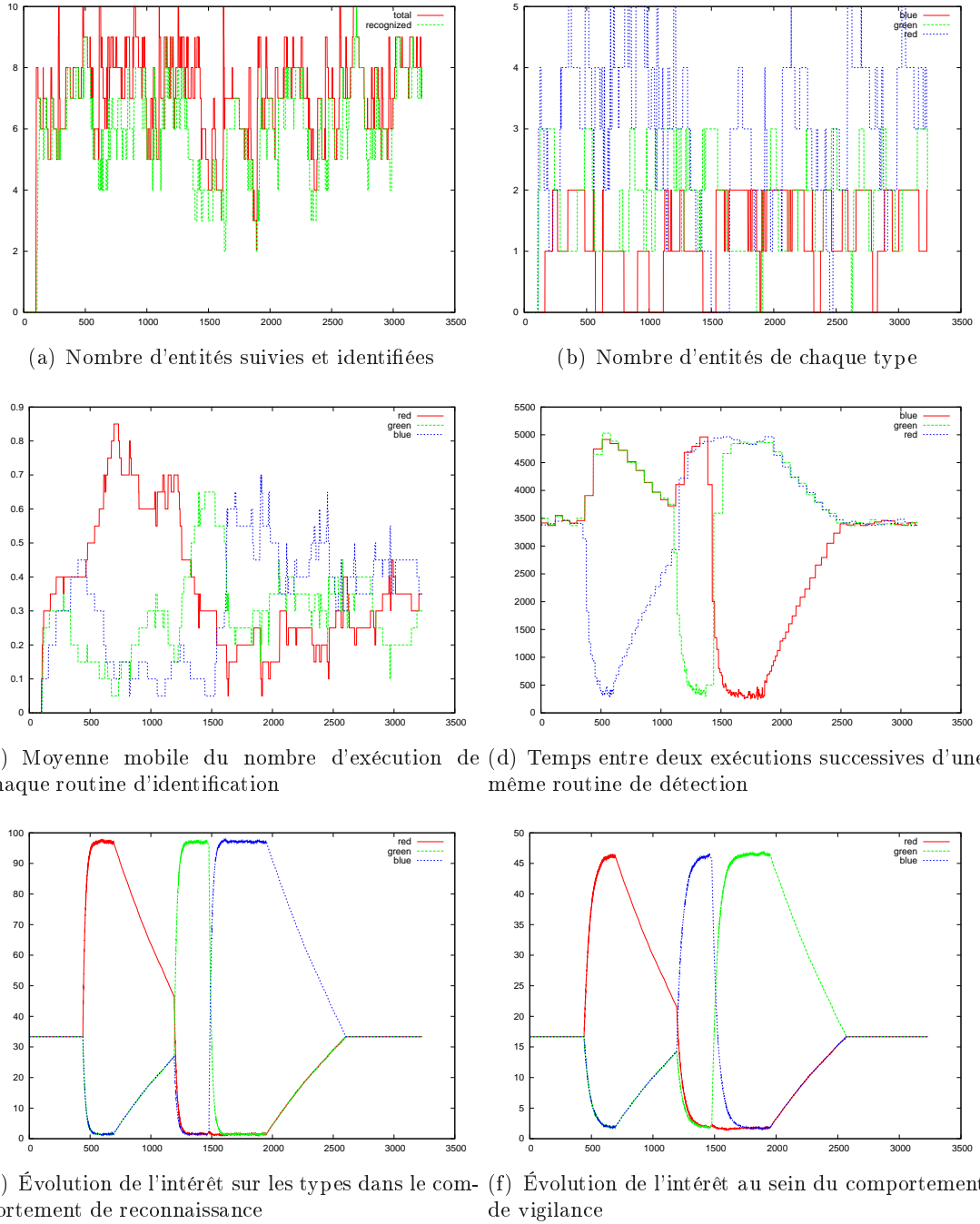


FIG. 6.16 – Paramètres du système sous influence du mécanisme de contrôle avec une limitation à 10 images par seconde.

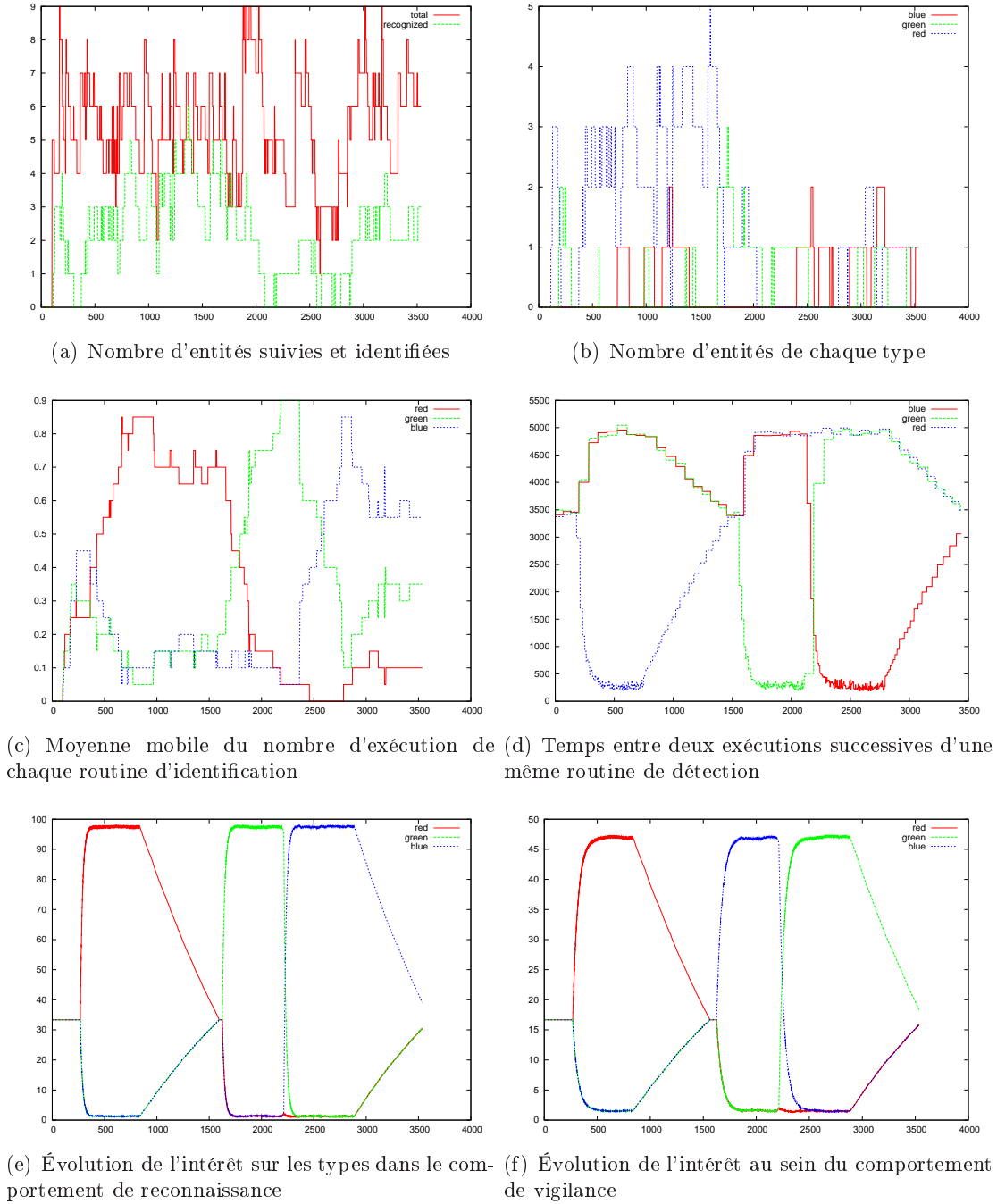


FIG. 6.17 – Paramètres du système sous influence du mécanisme de contrôle avec une limitation à 15 images par seconde.

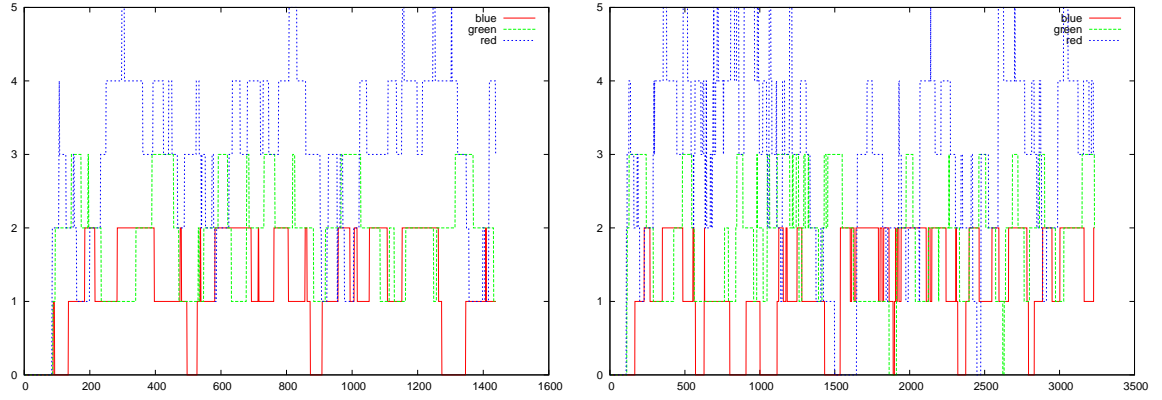


FIG. 6.18 – Nombre d’entités de chaque type sans et avec contrôle attentionnel pour une limitation à 15 images par secondes.

est néanmoins possible d’observer des différences par rapport au fonctionnement du système en l’absence de contrôle. Au niveau des variations du nombre d’entités de chaque type dans la représentation notamment (figure 6.18). Nous pouvons observer une diminution des entités de type **red** lorsque l’attention du système est portée sur les entités de type **green** puis **blue** sans laisser le temps à l’atténuation de faire son effet. L’effet du mécanisme de contrôle se manifeste au contraire notablement lorsque la limitation de la capacité de traitement passe à 15 images par secondes (figure 6.17), au niveau de l’évolution du nombre d’entités de chaque type l’effet est beaucoup plus marqué. On peut observer un pic pour chacun des types connus lorsque l’attention du système leur est portée (figure 6.17(b)). De même, les effets sur la moyenne mobile du nombre d’exécution des routines permettant d’identifier ces différents types son beaucoup plus marqués donnant lieu à l’observation de trois phases distinctes dans le fonctionnement du système (figure 6.17(c)). Pour chacune des différentes phases de contrôle, le nombre d’exécution des classifieurs correspondant au type recherché est en nette augmentation par rapport au nombre d’exécution des classifieurs associés aux autres types connus. Sur la fin de la session d’exécution limitée à 10 images par seconde, il est possible d’observer un retour à la normale du système du fait du mécanisme d’atténuation. Ceci ramène l’équilibre entre les différents traitements comme on peut le remarquer sur le moyenne mobile de l’exécution des différentes routines d’identification (figure 6.17(c)). Cela se note également sur le graphique décrivant le délai d’exécution entre deux routines de détection d’un même type d’entité 6.17(d)).

L’intérêt du mécanisme de contrôle attentionnel peut s’observer lorsque l’on compare les courbes représentant le nombre d’entités de chaque type présentes dans la représentation (figure 6.19). Les courbes de gauche présentent ces données en l’absence de contrôle du système de perception tandis que les courbes de droite présentent les mêmes données lorsque le contrôle est activé. Dans le premier cas, les ressources disponibles étant insuffisantes par rapport aux traitements nécessaires afin de représenter la totalité de l’environnement seules quelques entités de chaque type sont présentes en mémoire. Le nombre d’entités de chaque type est relativement équilibré et toutes ne sont bien évidemment pas détectées. Dans le second cas au contraire le nombre d’entités de chaque type est moins équilibré et dépend du type sur lequel l’attention du mécanisme de contrôle est porté. Ainsi, lorsque l’attention est

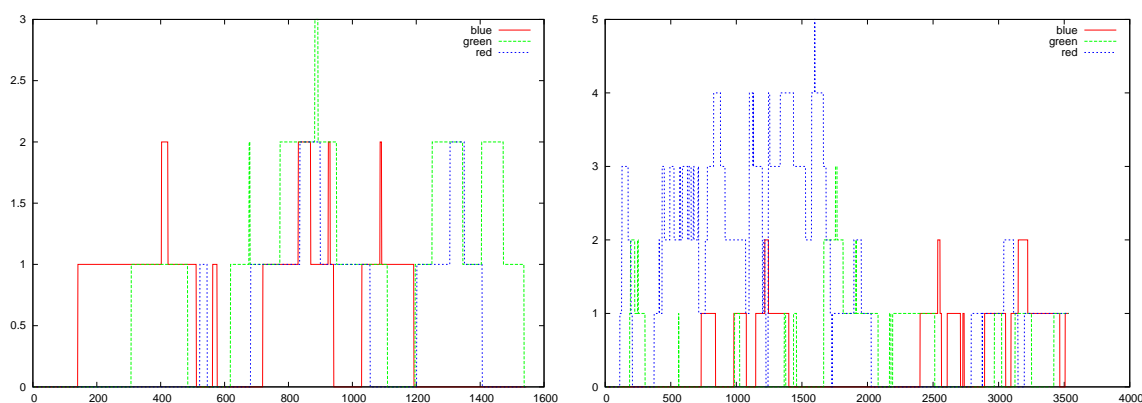


FIG. 6.19 – Nombre d’entités de chaque type sans et avec contrôle attentionnel pour une limitation à 15 images par secondes.

sur les entités de type **red** (première phase de contrôle) on peut observer un nombre d’entités de ce type plus important qu’en son absence.

L’influence du mécanisme attentionnel à proprement parler s’observe finalement sur les courbes représentant la moyenne mobile du nombre d’exécution des différentes routines d’identification et la variation du délai entre deux exécution d’une même routine de détection. Dans le premier cas, lorsque l’attention se porte sur un type d’entité particulier, le nombre d’exécution des routines d’identification de ce type d’entité augmente notablement. Dans le second, le fait de s’intéresser à un type d’entité particulier diminue le temps entre deux exécution de la routine de détection associée à ce type d’entité (*i.e.* augmentation du nombre de détections pendant une durée déterminée).

On peut également observer la façon dont les différentes routines se partagent une quantité fixée d’attention sur les courbes présentant l’évolution de l’intérêt au sein de comportement de reconnaissance et de vigilance (figures 6.16(e), 6.16(f), 6.17(e) et 6.17(f)). Lorsque l’attention est portée sur un type donné d’entités, la quantité d’intérêt disponible pour les autres types d’entités diminue de façon proportionnelle. De même lorsque l’on change l’objet de l’attention du mécanisme de contrôle, on peut observer une diminution importante de l’intérêt existant jusque là allant même jusqu’à faire disparaître l’effet de l’atténuation « naturelle » en faveur d’un retournement quasiment instantané.

Il est possible de contrôler ces variations au travers de deux paramètres :

- ▷ le pas d’incrémentation de l’intérêt lorsque par exemple un type particulier d’entité est recherché ;
- ▷ la vitesse d’atténuation définie en pourcentage de la quantité totale d’intérêt par seconde.

La figure 6.20 présente les variations d’intérêt pour le comportement de vigilance avec différents paramètres d’atténuation lors d’une session d’exécution de 30 secondes. Au cours de cette session, l’attention du système de perception est portée sur les entités de type **red** pendant environ 15 secondes. Les différentes courbes présentent des atténuation de 5%, 10%

et 50% par seconde (de haut en bas et de gauche à droite) pour un incrément de valeur 2 et une attention totale de 50. On peut observer que ce paramètre permet de contrôler la capacité à l'oubli du mécanisme de contrôle. Malgré tout il est important de considérer des valeurs en rapport avec l'incrément et le nombre d'incrémentations autorisées par seconde. Ceci peut se constater sur la courbe présentant un taux d'atténuation de 50%. Sur celle-ci, l'oubli est si rapide que le mécanisme attentionnel en perd son contrôle sur la sélection des routines de perception, les variations d'intérêt nécessitant quelques instants pour se répercuter sur le fonctionnement du système de perception.

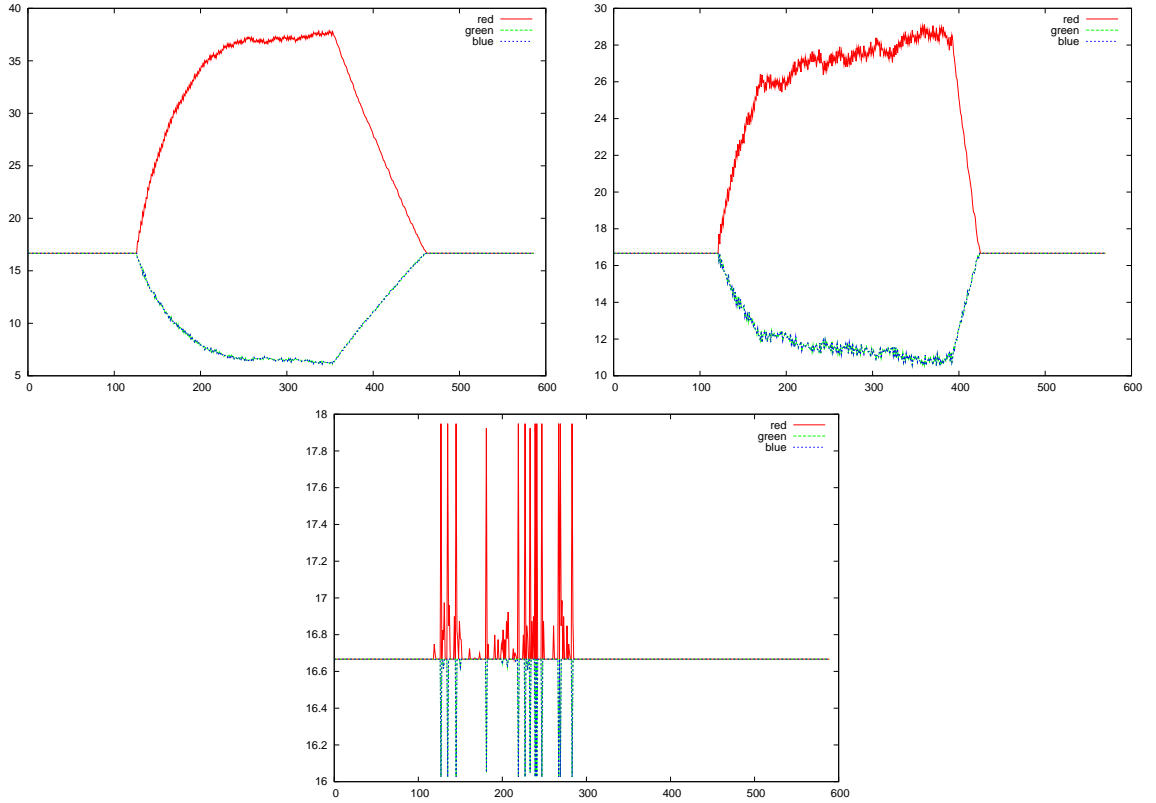


FIG. 6.20 – Évolution de l'intérêt au sein du comportement de vigilance pour différents taux d'atténuation (5%, 10% et 50% par seconde).

6.4.4 Bilan

Dans cette section nous nous sommes intéressés au système de perception visuelle au travers de plusieurs expérimentations sur des vidéos afin de mettre en évidence ses propriétés d'adaptation.

La première expérimentation nous a permis de mettre en évidence le fonctionnement du système de perception en l'absence de tout mécanisme de contrôle attentionnel. La mise en place d'une limite au niveau du temps de calcul imparti nous a permis de maintenir une cohérence plus ou moins forte entre l'environnement et sa représentation au prix de la

complétion de cette même représentation.

Au travers de la seconde expérimentation nous avons mis en évidence l'effet du mécanisme de contrôle attentionnel sur le contenu de la représentation. Nous avons également décrit son fonctionnement au sein de différents comportements (reconnaissance et vigilance) et nous nous sommes intéressés au problème du paramétrage du mécanisme attentionnel. Afin d'être exploitable par le processus de prise de décision, la représentation doit être fortement cohérente avec l'environnement, néanmoins il est nécessaire que celle-ci contienne un maximum d'informations. Pour cela le mécanisme de contrôle attentionnel permet de ne pas se limiter à un simple compromis entre temps de calcul et contenu de la représentation mais autorise une influence sur les traitements effectués par le système de perception visuelle, permettant ainsi de sélectionner indirectement les traitements qui sont effectués lorsque les ressources sont insuffisantes.

6.5 Conclusion

Dans ce chapitre nous avons tout d'abord présenté l'application de notre contribution à un cas réel : la visite guidée d'un aquarium marin. Pour cela nous avons présenté les différents comportements et données de discours définis dans le cadre du projet ANR SIRENE ainsi que des éléments sur le paramétrage du système de perception visuelle. Nous avons noté la difficulté de l'identification de façon « certaine » des différents poissons présents dans l'aquarium.

Nous avons ensuite présenté des expérimentations concernant les deux aspects de notre proposition : la visite guidée et le système de perception visuelle. Ces expérimentations nous ont permis de mettre en évidence les propriétés d'adaptation du système au niveau des explications en fonction des événements se produisant dans l'environnement et des connaissances *a priori* définies par les auteurs. Elles nous ont également permis de mettre en avant les propriétés du mécanisme de contrôle attentionnel.

Dans le cadre des expérimentations sur la visite guidée, nous avons explicité le fonctionnement du système et la gestion des interruptions à partir d'une évolution contrôlée de l'environnement. Nous nous sommes ensuite intéressés au problème du choix des explications en comparant différentes configurations d'experts, ici encore avec une évolution du contenu de la représentation de l'environnement définie par avance. Ceci nous a permis de mettre en avant l'influence des différents experts sur le déroulement de la visite guidée.

En ce qui concerne le système de perception visuelle, nous avons mené des expérimentations sur des vidéos représentant des scènes synthétiques afin de nous détacher des problèmes spécifiques au traitement d'images et aux difficultés introduites par l'environnement aquatique de notre application. Ceci nous a permis de mettre en évidence le fonctionnement du système avec et sans mécanisme de contrôle attentionnel en jouant à chaque fois sur la limitation des ressources disponibles. Nous avons ainsi pu mettre en avant les capacités d'adaptation offertes par l'architecture que nous proposons et les mécanismes d'interaction possibles entre processus de prise de décision et perception visuelle.

Partie V

— Bilan et perspectives —

Chapitre 7

Conclusion générale

7.1 Bilan

Problématique

Nous avons abordé le problème de l'utilisation d'un acteur virtuel autonome – le guide virtuel – pour décrire des objets dynamiques dans un environnement réel. Au delà des problèmes classiques de la réalité augmentée et des agents de présentation, nous nous sommes intéressés aux problèmes spécifiques introduits par la nature dynamique de l'environnement à décrire. Notre ambition était de proposer une solution selon les deux aspects principaux de ces problèmes : la description du réel (la visite guidée mise en œuvre par le guide virtuel) et la perception du réel (la perception en temps réel de l'environnement que l'on cherche à décrire). Nous avons abordé le premier aspect du point de vue de la structuration d'un discours, mettant de côté les problèmes d'animation et/ou de gestion des différentes modalités de communication. Le problème de la perception du réel a été considéré du point de vue de la perception visuelle par le biais de la vision par ordinateur (utilisation de caméras vidéo).

Du fait de la nature dynamique de l'environnement à décrire, la structure de la visite guidée ne peut être déterminée à l'avance : elle doit être construite au fur et à mesure de son déroulement. Ceci doit se faire en fonction du contexte (utilisateur, environnement, historique, guide, *etc*) et de contraintes relatives à sa structure globale (déroulement souhaité par les auteurs, les sujets qui doivent être abordés) et locale (cohérence des explications). L'utilisation d'un personnage pour incarner le guide virtuel au sein de l'environnement qu'il décrit implique de pouvoir le doter d'un comportement crédible dans cet environnement. Ici encore, l'aspect dynamique de l'environnement impose de pouvoir adapter ce comportement aux différents événements pouvant se produire.

La description de l'environnement est un processus qui dure dans le temps et nécessite de pouvoir connaître l'état de l'environnement à tout instant. Déterminer cet état nécessite la mise en œuvre de traitements coûteux qu'il n'est pas envisageable d'effectuer au moment où les informations qu'ils permettent d'obtenir sont requises. Il est donc indispensable de construire une représentation de cet état de l'environnement afin que le processus de prise de décision du

guide virtuel puisse l'interroger à tout instant. La construction d'une représentation complète de l'environnement et sa mise à jour en temps réel n'est néanmoins pas envisageable étant donnée la complexité des traitements requis et les ressources computationnelles limitées. De ce fait il faut être en mesure de sélectionner les traitements appropriés afin d'adapter le contenu d'une représentation partielle de l'environnement aux besoins de la visite guidée et du comportement du guide virtuel.

Proposition et résultats

Nous avons proposé un guide virtuel autonome dont l'architecture se décompose en trois modules (figure 7.1) : la prise de décision, la perception visuelle et l'animation. L'interaction de ces trois modules de façon asynchrone permet au système complet de développer une visite guidée tout en s'adaptant à un environnement dynamique peu ou pas prévisible.

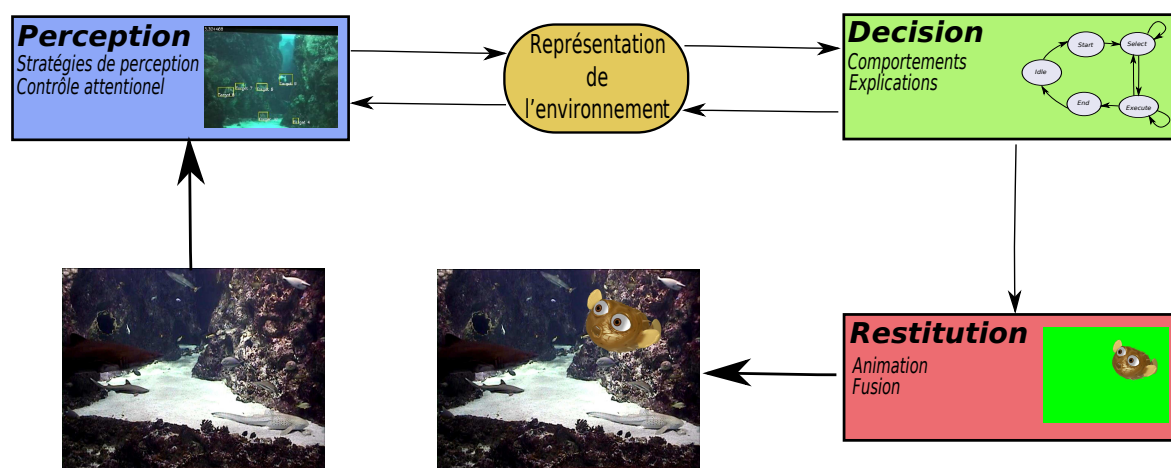


FIG. 7.1 – Vue d'ensemble de l'architecture proposée.

Nous considérons la visite guidée comme l'évolution conjointe du comportement du guide virtuel et des explications qu'il fournit au visiteur. Un graphe d'explication structure un ensemble d'éléments de discours atomiques relatifs à sujet donné. Chacun de ces éléments permet d'abstraire, sous la forme d'un script d'animation, une unité de discours indivisible pouvant faire appel à différentes modalités (parole, gestes, expression, *etc*). L'exécution d'un graphe d'explication (processus d'explication) conduit à l'exécution d'un ensemble d'éléments (exécution de scripts d'animation) correspondant à un chemin au sein de ce graphe. Le comportement du guide virtuel est sous la responsabilité d'un automate hiérarchique. Cet automate est constitué d'états composites et élémentaires, les état composites intégrant la notion d'interruption. Au cours de la visite guidée, ces interruptions entraînent une réévaluation du sujet courant pouvant conduire au choix d'un nouveau sujet à traiter. Le passage d'un sujet à un autre de manière cohérente est rendu possible grâce à l'utilisation de transitions sélectionnées en fonction de la situation. Le choix du sujet à traiter se fait de manière discrète par le vote d'un ensemble d'experts. Chaque expert définit un point de vue particulier sur le processus de sélection par rapport au contexte courant de la visite (historique, environnement, *etc*).

Le système de perception visuelle est centré sur la notion de représentation de l'environnement. Il consiste en un ensemble de routines de perception qui coopèrent à la construction et à la mise à jour d'une représentation commune. Cette représentation est constituée d'un ensemble d'entités visuelles décrites par des propriétés de types variés (visuelles, géométriques, spatiales, *etc*), leur valeur étant le résultat de l'exécution d'une routine particulière. Certaines routines mettent en œuvre des stratégies de perception en sélectionnant et paramétrant les traitements qu'elles déclenchent en fonction du contexte. Ces traitements peuvent consister en le calcul d'une propriété par une autre routine ou l'exécution de traitements de bas niveau (filtres appliqués aux images, *etc*). Les différentes stratégies sont adaptées en fonction des besoins du processus de prise de décision du guide virtuel par rapport au contenu courant de la représentation de l'environnement et aux requêtes effectuées par le processus de prise de décision sur cette même représentation. La combinaison de ces deux influences permet indirectement d'adapter les informations contenues dans la représentation à la fois aux besoins du processus de prise de décision et à l'état de l'environnement réel. L'influence du processus de prise de décision se fait par l'intermédiaire d'objets d'intérêt. La nature de ces différents objets dépend de la stratégie mise en œuvre. Nous avons proposé trois stratégies (vigilance, suivi et reconnaissance) que nous considérons comme nécessaires dans le cadre de la représentation d'un environnement réel dynamique. Celles-ci permettent la détection, le suivi et l'identification des entités visibles par les caméras vidéo.

Les travaux menés dans cette thèse ont permis la réalisation d'une application dans le cadre de la visite guidée d'un bassin de l'aquarium marin Océanopolis. Celle-ci nous a permis de valider notre approche sur un cas réel d'environnement dynamique. Des expérimentations concernant la visite guidée nous ont permis d'explicitier le fonctionnement des mécanismes d'interruption que nous avons proposés et de mettre en évidence le processus de sélection de sujet par l'exécution de différentes combinaisons d'experts avec une évolution connue de la représentation de l'environnement. Les expérimentations relatives au système de perception nous ont permis d'étudier le fonctionnement du mécanisme de contrôle attentionnel en mettant de côté la difficulté du problème de la reconnaissance liée à l'application. Nous y avons décrit l'influence du mécanisme de contrôle sur les différentes stratégies de perception ainsi que sur le contenu de la représentation de l'environnement.

Discussion

Dans cette thèse nous avons proposé une solution quant à la description d'un environnement réel dynamique par un guide virtuel autonome. Nous avons répondu aux différentes questions posées par cette problématique suivant deux aspects relativement éloignés (la description du réel et la perception du réel) et proposé un mécanisme d'interaction simple entre ces aspects.

La solution proposée quant à la description du réel (la visite guidée mise en œuvre par le guide virtuel) met en avant les difficultés introduites par l'aspect dynamique de l'environnement. Le mécanisme d'interruption proposé ainsi que la notion d'explication relativement modulaire permet au guide virtuel d'adapter son discours de façon appropriée en fonction du contexte. Cette notion d'explication simple présente également un avantage du point de vue des auteurs en ne reposant sur aucune théorie linguistique complexe mais simplement sur un découpage précis du contenu pouvant être présenté. Le choix à l'aide d'un ensemble d'experts des sujets devant être traités offre un contrôle précieux à ces mêmes

auteurs malgré l'imprévisibilité de l'environnement à décrire. Ce mécanisme offre également une grande modularité lors de la création d'une visite guidée.

Le système de perception visuelle présente quant à lui une architecture générique que nous avons adaptée aux besoins de la description d'un environnement réel dynamique. Cette architecture centrée sur une représentation partielle de l'environnement peut être facilement étendue. La caractéristique principale du système que nous avons proposé réside dans sa capacité de focalisation en fonction à la fois du contenu de la représentation de l'environnement et des besoins du processus de prise de décision. La notion de stratégie de perception permet d'effectuer finement cette focalisation suivant les spécificités des différents aspects de la perception (suivi, reconnaissance, *etc*), chacun présentant des difficultés particulières. Le mécanisme d'interaction que nous avons proposé permet une certaine indépendance entre le processus de prise de décision et le système de perception visuelle grâce à l'utilisation de la représentation de l'environnement en tant qu'interface. Ceci permet de faciliter l'utilisation de chaque partie dans d'autres applications.

En ce qui concerne l'application à la visite guidée d'un aquarium marin, notre proposition a été développée de paire avec cette application et est de ce fait relativement adaptée à la situation. Néanmoins, la réalisation de cette application nous a permis de mettre en évidence un certain nombre de difficultés quant à l'identification des différentes espèces de poissons présentes dans l'aquarium. Reste également à voir comment notre proposition pourrait être étendue à d'autres types d'applications muséographiques (système mobile où le guide accompagne le visiteur lors de ses déplacements dans le musée) ou non (explications/commentaires sur des simulations numériques ou multi-agents dans lesquelles des événements imprévisibles peuvent se produire).

7.2 Perspectives

La réalisation de ce travail de thèse a ouvert plusieurs axes de réflexion et offre différentes perspectives.

Sur l'application

Parmis les résultats que nous avons présentés il manque une évaluation précise auprès du public d'Océanopolis. Celle-ci n'a pu être réalisée par manque de temps. Il nous a néanmoins été possible de présenter brièvement l'application lors de la « Nuit des Chercheurs ». Le public rencontré lors de cet événement n'est bien évidemment pas le même que le public quotidien d'Océanopolis mais certaines de ses remarques sont à considérer.

La remarque principale concerne l'interactivité, la plupart des visiteurs souhaiterait pouvoir communiquer/interagir avec le guide virtuel. Bien qu'un système complètement interactif à la manière d'un kiosque d'information ne soit pas l'objectif premier de l'idée du guide virtuel, certaines possibilités sont à envisager. Il pourrait en effet être intéressant de tirer parti de cette notion d'interactivité afin de lever des ambiguïtés lors de l'identification des différents poissons. Ceci pourrait se faire sous la forme d'un jeu demandant au visiteur

de choisir parmi plusieurs possibilités. Cela nécessiterait des capacités d'introspection de la part du système de perception visuelle, qu'il soit capable de connaître les différentes options qui s'offrent à lui en terme d'identification. De plus, l'ajout d'interactions avec le visiteur permettrait d'augmenter la crédibilité du guide virtuel.

Sur la description du réel

Au delà des problèmes d'ingénierie tels que la réalisation d'outils auteurs permettant de décrire les explications et le comportement du guide virtuel plus simplement, certains points pourraient être poussés plus avant.

L'aspect pédagogique tout d'abord avec par exemple la prise en compte du niveau de connaissances du visiteur, d'une intention dans le discours du guide virtuel et/ou la mise en place de visites guidées sur un thème spécifique. Ceci pourrait être fait en définissant par exemple de nouveaux experts plus complexes que ceux que nous avons présentés. Concernant ces derniers, il serait intéressant de développer les expérimentations que nous avons menées afin d'identifier plus précisément l'impact de chacun d'entre eux sur le déroulement global de la visite guidée.

Le système de transitions entre sujets que nous avons mis en place mériterait également des améliorations. La notion large de transition – au sens où celles-ci incluent les introductions et conclusions relatives aux différents sujets – que nous avons proposée présente en l'état un certain nombre de limitations. La spécificité des transitions ainsi que les différentes combinaisons envisageables entre ces transitions (conclusion suivie d'une introduction par exemple) sont pour l'instant à la charge du comportement exécutant les explications. Ceci est trop spécifique et nécessite un lien étroit entre la définition du comportement et l'élaboration du contenu pouvant être présenté. Une amélioration pourrait consister en la prise en compte des caractéristiques de spécificité (niveau d'adaptation à la situation courante) et de type (introduction, conclusion ou changement de sujet par exemple) des différents éléments de transition. L'ajout de telles informations pourrait permettre la mise en place de mécanismes de combinaison d'éléments de transition simples afin de construire une transition unique adaptée au mieux à la situation.

Du point de vue de la présentation des informations il serait intéressant d'étendre le système actuel en intégrant les problématiques « classiques » des agents de présentation. Ainsi, les scripts d'animation pourraient par exemple être remplacés par un système de gestion et de choix des modalités. Cela nécessiterait néanmoins la définition d'un grand nombre de connaissances supplémentaires qui ne sont pas nécessaires à l'heure actuelle.

Un dernier aspect enfin concerne la gestion des supports du discours lors du déroulement de la visite guidée. Ici nous n'avons considéré que les entités visuelles en tant que support dynamique du discours du guide virtuel. Il pourrait être intéressant de réifier cette notion de support d'explication afin par exemple d'être capable de choisir dynamiquement un support virtuel (une image ou vidéo) ou réel (une entité visuelle) suivant l'état de l'environnement et le contexte de la visite guidée.

Sur la perception du réel

Ici encore, mises à part les perspectives concernant le logiciel développé durant cette thèse telles que la mise en place de procédure d'apprentissage automatisées pour la reconnaissance ou l'ajout de nouveaux traitements au répertoire existant, plusieurs aspects du système de perception visuelle que nous avons proposé pourrait se voir étendus.

Du point de vue de la reconnaissance, l'utilisation d'une approche multi-experts pourrait s'avérer pertinente. La définition de processus de classification spécifique suivant différentes zones du champ visuel ou poses des entités à percevoir permettrait certainement d'améliorer les résultats obtenus actuellement. Au delà de l'amélioration des performances du système de reconnaissance actuel, il serait intéressant d'étendre ses capacités pour par exemple permettre l'identification de situation particulière et/ou d'interactions entre les différentes entités visuelles. Ceci pourrait être effectué en ajoutant de nouvelles stratégies et routines de perception spécialisées et permettrait d'adapter plus finement le discours à l'état de l'environnement perçu.

La distribution des traitements mériterait également que l'on s'y attarde. Nous avons déjà abordé ce point dans la mise en œuvre actuelle. Néanmoins il serait intéressant de mener des expérimentations quant au gain réel apporté par la distribution des traitements au sein de différents processus. Ici plusieurs pistes sont envisageables comme le regroupement des traitements fortement liés au sein d'un même processus afin de minimiser les envois de messages (ce qui est fait actuellement) ou la duplication de routines de perception au sein de différents processus (parallélisation des traitements). Bien que ce problème soit clairement lié à notre mise en œuvre et l'architecture logicielle que nous utilisons, il nous semble que le système de perception visuelle que nous proposons se prête particulièrement bien à une telle étude.

Une autre perspective qui pourrait être explorée concerne l'apprentissage. Au delà de l'automatisation des procédures permettant l'ajout de nouveaux processus de classification, l'idée serait de permettre au système lui-même de construire ces processus en fonction de leur efficacité. Ce point n'a jusqu'à présent pas été abordé mais des pistes telles que l'utilisation d'un processus évolutionnaire sont envisageables.

L'interaction entre le processus de prise de décision et le système de perception visuelle par la représentation de l'environnement nécessiterait également des travaux supplémentaires. Les différentes requêtes et objets d'intérêt nécessitent des connaissances sur les stratégies de perception et le fonctionnement du système de perception visuelle. En l'état ces connaissances sont disséminées au sein de la définition des différentes requêtes possibles sur la représentation de l'environnement. Proposer une modélisation plus formelle de ces connaissances permettrait de rendre le mécanisme de contrôle attentionnel plus générique en automatisant la définition des requêtes et différents objets d'intérêt.

Enfin, un dernier point qu'il serait intéressant d'explorer serait de faire tendre le système de perception visuelle vers un système multi-agents. En dotant de capacité de perception et de prise de décision les différentes routines de perception (et plus seulement les comportements de perception), le système au complet pourrait être en mesure de s'adapter aux besoins du processus de prise de décision ainsi qu'aux ressources disponibles. Ceci nécessiterait bien évidemment de régler un certain nombre de problèmes complexes et notamment de revoir

sensiblement la notion d'objet d'intérêt. Néanmoins, l'architecture actuelle permet d'ores et déjà une telle extension.

Références bibliographiques

- [Abowd 97] Gregory D. Abowd, Christopher G. Atkeson, Jason Hong, Sue Long, Rob Kooper & Mike Pinkerton. *Cyberguide : A mobile context-aware tour guide*. Wireless Networks, vol. 3, pages 421–433, 1997.
- [Agre 95] Philip E. Agre & David Chapman. *Pengi : an implementation of a theory of activity*. pages 635–644, 1995.
- [Allen 01] B. Danette Allen, Gary Bishop & Greg Welch. *Tracking : Beyond 15 Minutes of Thought*. SIGGRAPH course, 2001.
- [Aloimonos 88] John Aloimonos, Isaac Weiss & Amit Bandyopadhyay. *Active Vision*. International Journal of Computer Vision, vol. 1, no. 4, pages 1573–1405, January 1988.
- [Aloimonos 90] John Aloimonos. *Purposive and qualitative active vision*. 10th International Conference on Pattern Recognition, vol. 1, pages 346–360, 16-21 June 1990.
- [Amendola 04] Haria Amendola, Federica Cena, Luca Console, Andrea Crecola, Cristina Gena, Anna Goy, Sonia Modeo, Monica Perrero, Ilaria Torre & Andrea Toso. *UbiquiTO : a Multi-Device Adaptive Guide*. In Mobile Human-Computer Interaction, 2004.
- [Andre 98] E. Andre, T. Rist & J. Mueller. *Employing AI methods to control the behavior of animated interface agents*, 1998.
- [André 88] E. André, G. Herzog & T. Rist. *On the simultaneous interpretation of real world images sequences to their natural language description : The system SOCCER*. In proceedings of the 8th ECAI, pages 449–454, 1988.
- [André 97] E. André, G. Herzog & T. Rist. *Generating multimedia presentations for RoboCup soccer games*. Lecture notes in computer science, vol. 1395, pages 200–215, 1997. Robocup Workshop 1997.

- [André 00a] Elisabeth André, Kim Binsted, Kumiko Tanaka-Ishii, Sean Luke, Gerd Herzog & Thomas Rist. *Three RoboCup Simulation League Commentator Systems*. AI Magazine, 2000.
- [André 00b] Elisabeth André. Handbook of natural language processing, chapitre The Generation of multimedia presentations, pages 305–328. CRC Press, 2000.
- [Arc] *Archimuse*. <http://www.archimuse.com>.
- [Arguin 03] Martin Arguin. Perception et réalité, chapitre L’attention sélective, pages 426–443. de boeck, 2003.
- [Atkinson 68] R.C. Atkinson & R.M. Shiffrin. *Human memory : A proposed system and its control processes*. The psychology of learning and motivation, vol. 8, 1968. In K.W. Spence and J.T. Spence (Eds.). London : Academic Press.
- [Augello 07] A. Augello, A. Santangelo, S. Sorce, G. Pilato, A. Gentile, A. Genco & S. Gaglio. *A Multimodal Interaction Guide for Pervasive Services Access*. In IEEE International Conference on Pervasive Services, pages 250–256, 15–20 July 2007.
- [Aylett 99] R.S. Aylett. *Narrative in Virtual Environments : Towards Emergent Narrative*. Rapport technique, Papers from the 1999 AAAI Fall Symposium, Technical report FS-99-01, 1999. AAAAI Press, Menlo Park, 1999 pp83-86, ISBN 1-57735-103-7.
- [Azuma 01] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier & Blair MacIntyre. *Recent Advances in Augmented Reality*. IEEE Computer Graphics and Applications, 2001.
- [Baccon 00] J.-C. Baccon, L. Hafemeister & Ph. Gaussier. *Computational model for a task directed attention mechanism without effective recognition of targets*. In ECV Workshop, Isle of Skye - Scotland, 2000.
- [Backer 03] Gerriet Backer & Bärbel Mertsching. *Two selection stages provide efficient object-based attentional control for dynamic vision*, 2003.
- [Baddeley 74] A.D. Baddeley & G.J. Hitch. *Working memory*. In g. bower, editeur, The psychology of learning and motivation, volume 3, pages 47–89. New York : Academic press, 1974.
- [Bajcsy 88] Ruzena Bajcsy. *Active Perception*. Proceedings of the IEEE, Special issue on Computer Vision, vol. 76, no. 8, August 1988.
- [Ballard 82] Dana H. Ballard & Christopher M. Brown. Computer vision. Prentice Hall, 1982.
- [Ballard 91] Dana H. Ballard. *Animate Vision*. Rapport technique 329, University of Rochester, 1991.

- [Ballard 95] D. Ballard & R. Rao. *A computational model of human vision based on visual routines*, 1995.
- [Baluja 95] Shumeet Baluja & Dean A. Pommerleau. *Using a Saliency Map for Active Spatial Selective Attention : Implementation & Initial Results*. In G. Tesauro, D. Touretzky & T. Leen, éditeurs, *Advances in Neural Information Processing Systems*, volume 7, pages 451–458. The MIT Press, 1995.
- [Bar 93] Moshe Bar & Shimon Ullman. *Spatial Context in Recognition*, 1993.
- [Bar 04] Moshe Bar. *Visual Objects in Context*. *Nature Neuroscience*, vol. 5, pages 617–629, August 2004.
- [Bateman 03] John A. Bateman. *Encyclopedia of library and information science*, chapitre Automatic Discourse Generation, pages 257–269. CRC Press, 2003.
- [Bearman 91] David Bearman. *Interactive and Hypermedia in Museums*. In *International Conference on Hypermedia and Interactivity in Museums*, 1991.
- [Bearman 95] David Bearman. *Hands-on : A 1995 Snapshot of the evolution of interactive multimedia*. In *Third International Conference on Hypermedia and Interactivity in Museums (ICHIM 95 / MCN 95)*, 1995.
- [Bellotti 02] F. Bellotti, C. Berta, A. de Gloria & M. Margarone. *User testing a hypermedia tour guide*. *Pervasive Computing, IEEE*, vol. 1, no. 2, pages 33–41, April–June 2002.
- [Bergamasco 01] M. Bergamasco, C. A. Avizzano, G. Di Pietro, F. Barbagli & A. Frisoli. *The museum of pure form*. *IEEE International workshop on robot and human interactive communication*, 2001.
- [Beskow 97a] Jonas Beskow. *Animation of Talking Agents*. In *Proceedings of AVSP'97, ESCA Workshop on Audio-Visual Speech Processing*, pages 149–152, Rhodes, Greece, September 1997.
- [Bins 05] J. Bins, T. List, R. Fisher & D. Tweed. *An intelligent and task-independent controller for video sequence analysis*. *IEEE International Workshop on Computational Architecture for Machine Perception*, 2005.
- [Binsted 98] K. Binsted. *Character design for soccer commentary*. In *Proceedings of the second RoboCup workshop*, pages 23–35, 1998.
- [Björk 98] S. Björk. *Making Guides Entertaining*. In *Workshop paper at ECAI Workshop on AI and Entertainment*, Brighton, UK, 1998.

- [Bollmann 97] Maik Bollmann, Rainer Hoischen & Barbel Mertsching. *Integration of Static and Dynamic Scene Features Guiding Visual Attention*. In DAGM-Symposium, pages 483–490, 1997.
- [Bordegoni 97] M. Bordegoni, G. Faconti, S. Feiner, M.T. Maybury, T. Rist, S. Ruggieri, P. Trahanias & M. Wilson. *A Standard Reference Model for Intelligent Multimedia Presentations Systems*. Computer Standards and Interfaces : The international journal on the development and applications of standards for computers, data communications and interfaces, vol. 18, pages 477–496, 1997.
- [Bordeux 99] Christophe Bordeaux, Ronan Boulic & Daniel Thalmann. *An Efficient and Flexible Perception Pipeline for Autonomous Agents*. In P. Brunet & R. Scopigno, éditeurs, Computer Graphics Forum (Eurographics '99), volume 18(3), pages 23–30. The Eurographics Association and Blackwell Publishers, 1999.
- [Boucher 99] Alain Boucher. *Un approche décentralisée et adaptative de la gestion d'information en vision. Application à l'interprétation d'images de cellules en mouvement*. PhD thesis, Université Joseph Fourier, Grenoble, January 1999.
- [Braun 03] Norbert Braun. *Storytelling in Collaborative Augmented Reality Environments*. In Science Press, éditeur, WSCG Short Papers, 3–7 February 2003.
- [Brill 94] Frank Brill. *Perception and Action in a Dynamic Three-Dimensional World*, 1994.
- [Brink 03] Ingar Brink. *The objects of attention*. In Proceedings of European Society for Philosophy and Psychology, Italy, Turin, July 2003.
- [Brooks 81] R. A. Brooks. *Symbolic Reasoning among 3D models and 2D images*. IEEE Journal of Robotics and Automation, vol. 2, no. 1, 1981.
- [Brooks 91a] R.A. Brooks. Architectures for intelligence, chapitre How to build complete creatures rather than isolated cognitive simulators, pages 225–239. Lawrence Erlbaum Associates, Hillsdale, NJ, 1991.
- [Brusilovsky 03] Peter Brusilovsky. *From Adaptive Hypermedia to the Adaptive Web3D*. In Mensch & Computer 2003 : Interaktion in Bewegung, pages 21–24, 2003.
- [Buche 05] Cedric Buche. *Un système tutoriel intelligent et adaptatif pour l'apprentissage de compétences en environnement virtuel de formation*. PhD thesis, Université de Bretagne Occidentale (UBO), Novembre 2005.

- [Burgard 98] W. Burgard, A.B. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, & S. Thrun. *The Interactive Museum Tour-Guide Robot*. In Proc. of the Fifteenth National Conference on Artificial Intelligence (AAAI-98), 1998.
- [Burgard 99] Wolfram Burgard, Armin B. Cremers, Dieter Fox, Dirk Hähnel, Gerhard Lakemeyer, Dirk Schulz, Walter Steiner & Sebastian Thrun. *Experiences with an interactive museum tour-guide robot*. Artificial Intelligence 00, May 1999.
- [Burges 98] Christopher J. C. Burges. *A Tutorial on Support Vector Machines for Pattern Recognition*. Data Mining and Knowledge Discovery, vol. 2, pages 121–167, 1998.
- [Cassell 94] Justine Cassell, Catherine Pelachaud, Norman Badler, Mark Steedman, Brett Achorn, Tripp Becket, Brett Douville, Scott Prevost & Matthew Stone. *Animated conversation : rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents*. Computer Graphics, vol. 28, no. Annual Conference Series, pages 413–420, 1994.
- [Cassell 01] Justine Cassell. *Embodied Conversational Agents - Representation and Intelligence in User Interfaces*. AI Magazine, vol. 22, no. 4, pages 67–83, 2001.
- [Cavanagh 99] P. Cavanagh. Mit encyclopedia of cognitive science, chapitre Top-down processing in vision, pages 844–845. MIT Press, 1999.
- [Cavanagh 01] Patrick Cavanagh, Angela T. Labianca & M. Thornton. *Attention-based visual routines : sprites*. Cognition, vol. 80, pages 47–60, 2001.
- [Cavanagh 04] Patrick Cavanagh. Attention routines and the architecture of selection, chapitre 2, pages 13–28. Guilford Press, 2004.
- [Cavanagh 05] Patrick Cavanagh & George A. Alvarez. *Tracking multiple targets with multifocal attention*. TRENDS in Cognitive Sciences, vol. 9, no. 7, pages 349–354, July 2005.
- [Cavazza 02] Marc Cavazza, Fred Charles, & Steven J. Mead. *Character-Based Interactive Storytelling*. IEEE INTELLIGENT SYSTEMS, pages 17–24, 2002.
- [Chapman 90] David Chapman. *Vision, Instruction, and Action*. Rapport technique AITR-1204, 1990.
- [Chetverikov 95] D. Chetverikov. *Is Computer Vision Possible ?* In Workshop Vision Milestones'95, Vorau, Austria, 1995. Austrian Society for Artificial Intelligence.
- [Chevallier 05] Sylvain Chevallier & Hélène Paugam-Moisy. *Modélisation de processus attentionnels dans la perception multimodale d'un robot virtuel*, 2005.

- [Cheverst 00] Keith Cheverst, Nidel Davies, Keith Mitchell & Adrian Friday. *Experiences of Developing and Deploying a Context-Aware Tourist Guide : The GUIDE Project*. In Sixth Annual International Conference on Mobile Computing and Networking, pages 20–31. ACM Press, 2000.
- [Chittaro 03] Luca Chittaro, Roberto Ranon & Lucio Ieronutti. *Guiding Visitors of Web3D Worlds through Automatically Generated Tours*. In ACM Press, éditeur, 8th International Conference on 3D Web Technology, 2003.
- [Choi 03] Sang-Bok Choi, Sang-Woo Ban & Minhoo Lee. *Biologically Motivated Visual Attention System Using Bottom-up Saliency Map and Top-down Inhibition*. Neural Information Processing-Letters and Reviews, vol. 2, no. 1, January 2003.
- [Chou 04] Li-Der Chou, Ching-Chao Lee, Ming-Yu Lee & Chun-Yen Chang. *A Tour Guide System for Mobile Learning in Museums*. In Proceedings of the 2nd IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE'04), page 195. IEEE Computer Society, 2004.
- [Chun 00] Marvin M. Chun. *Contextual cueing of visual attention*. Trends in cognitive sciences, vol. 4, no. 5, pages 170–178, May 2000.
- [Churchland 94] Patricia Churchland, V. S. Ramachandran & Terrance Sejnowski. Large-scale neuronal theories of the brain, chapitre A critique of pure vision. MIT Press, 1994.
- [Colineau 03] Nathalie Colineau & Cécile Paris. *La génération de documents multimédia*. In Cognito, vol. 1, no. 2, pages 1–22, 2003.
- [Concepcion 96] Vicente Concepcion & Harry Wechsler. *Detection and localization of objects in time varying imagery using attention, representation and memory pyramids*. Pattern Recognition, vol. 29, no. 9, pages 1543–1557, 1996.
- [Connor 04] Charles E. Connor, Howard E. Egeth & Steven Yantis. *Visual attention : bottom-up versus top-down dispatch*. Current Biology, vol. 14, pages R850–R852, 2004.
- [Corkill 91] Daniel D. Corkill. *Blackboard Systems*. AI Expert, vol. 6, no. 9, pages 40–47, September 1991.
- [Courty 03b] Nicolas Courty, Eric Marchand & Bruno Arnaldi. *A New Application for Saliency Maps : Synthetic Vision of Autonomous Actors*. In IEEE Int. Conf. on Image Processing, ICIP'03 Barcelona, Spain, September 2003.
- [Cowan 01] N. Cowan. *The magical number 4 in short-term memory : A reconsideration of mental storage capacity*. Behavioral and Brain Sciences, vol. 24, pages 87–185, 2001.

- [Craik 72] F. I. M. Craik & R. S. Lockhart. *Levels of Processing : A Framework for Memory Research*. Journal of Verbal Learning and Verbal Behavior, vol. 11, pages 671–684, 1972.
- [Crowley 94] J. L. Crowley, J. M. Bedrune, M. Bekker & M. Schneider. *Integration and Control of reactive visual processes*. In Proceedings of the first European Conference on Computer Vision, ECCV-94, pages 47–58, 1994.
- [Culhane 92] Sean M. Culhane & John K. Tsotsos. *An Attentional Prototype for Early Vision*. In G. Sandini, éditeur, Proceedings of the Second European Conference on Computer Vision, pages 551–560, Santa Margherita Ligure, Italy, May 1992. Springer Verlag.
- [Davies 98] N. Davies, K. Mitchell, K. Cheverst & G. Blair. *Developing a Context Sensitive Tourist Guide*, 1998.
- [de Almeida 03] Pilar de Almeida & Shigeki Yokoi. *Interactive character as a virtual tour guide to an online museum exhibition*. Museums and the web 2003, 2003.
- [de Carolis 02] B. de Carolis, V. Carofiglio & C. Pelachaud. *From discourse plans to believable behavior generation*, 2002.
- [de Fockert 01] Jan W. de Fockert, Geraint Rees, Christopher D. Frith & Nilli Lavie. *The Role of Working Memory in Visual Selective Attention*. Science, vol. 291, pages 1803–1806, March 2001.
- [Deco 02] Gustavo Deco, Olga Pollatos & Josef Zihl. *The time course of selective visual attention : theory and experiments*. Vision Research, vol. 42, pages 2925–2945, 2002.
- [Delorme 03] André Delorme & Michelangelo Flückiger. Perception et réalité, chapitre La problématique et les courants théoriques, pages 20–40. de boeck, 2003.
- [Deshayes 04] Sophie Deshayes. *L’usage des supports mobiles au musée, des audioguides classiques au multimédia nomade*. In ICHIM04, 2004.
- [Devillers 02] F. Devillers, S. Donikian, F. Lamarche & JF. Taille. *A programming environment for behavioural animation*. The Journal of Visualization and Computer Animation, 2002.
- [Dinis 08] A. Dinis, N. Fies, S. Otmane, M. Mallem, A. Nisan, J.M. Boï, C. Noel & C. Viala. *Digital Ocean : A national project for the creation and distribution of multimedia content for underwater sites*. In VSMM’08, Proceedings of the 14th international Conference on Virtual Systems and Multimedia, pages 389–396, 2008.
- [Don 91] Abbe Don, Tim Oren & Brenda Laurel. *Guide 3.0*. In Proceedings of the SIGCHI conference on Human factors in computing systems : Reaching through technology, 1991.

- [Donikian 01] S. Donikian. *HPTS : a behaviour modelling language for autonomous agents*. In Proceedings of AGENTS'01, pages 401–408, 2001.
- [Doyle 97a] P. Doyle & B. Hayes-Roth. *Guided exploration of virtual worlds*. Rapport technique KSL 97-04, Knowledge System Laboratory, Stanford University, 1997.
- [Doyle 99] P. Doyle & K. Isbister. *Touring machines : Guide agents for sharing stories about digital places*, 1999.
- [Draper 89] B. Draper, Robert Collins, J. Brolio, A. Hanson & E. Riseman. *The Schema System*. The International Journal of Computer Vision, vol. 2, no. 3, pages 209–250, January 1989.
- [Duchowski 00] Andrew Duchowski & Roel Vertegaal. *Eye-based interaction in graphical systems theory & practice*. SIGGRAPH Course 05, July 2000.
- [Economou 98] Maria Economou. *The evaluation of museum multimedia applications : lessons from research*. Museum Management and Curatorship, vol. 17, no. 2, pages 173–187, June 1998.
- [Edelman 89] S. Edelman & T. Poggio. *Representations in High-Level Vision : Reassessing the Inverse Optics Paradigm*. In Proceedings of the DARPA IUS Workshop, pages 944–949, 1989.
- [Edgington 04] Duane R. Edgington, Danelle E. Cline, R.E. Sherlock, Dirk Walther & Christof Koch. *Detecting and Tracking Animals in Underwater Video*. In CVPR'04, 2004.
- [Edgington 05] Duane R. Edgington, Ishbel Kerkez, Danelle E. Cline, Dorothy Oliver, Marc'Aurelio Ranzato & Pietro Perona. *Detecting, Tracking and Classifying Animals in Underwater Video*. In IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, California, 2005.
- [Fabiani 96] Patrick J. Fabiani. *Dynamics of Belief and Strategy of Perception*. In W. Wahlster, editeur, ECAI96. 12th European Conference on Artificial Intelligence. John Wiley & Sons, Ltd, 1996.
- [Fantoni 03] Silvia Filippini Fantoni. *MUSEUMS WITH A PERSONAL TOUCH*. J., Cappellini, V. and Stanke, G.(eds.), EVA London, 2003.
- [Findlay 01] J. M. Findlay & Gilchrist. Vision and attention, chapitre Visual attention : the active vision perspective, pages 85–106. Springer Verlag, 2001.
- [Franklin 96] Stan Franklin & Art Graesser. *Is it an Agent or just a program ? A taxonomy for autonomous agents*. In Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages, 1996.

- [Franz 03] Klemens Franz & Alexander K. Nischelwitzer. *Adaptive Digital Storytelling : A Concept for Narrative Structures and Digital Storytelling build on Basic Storytelling Principles, Adaptive Story Schemas and Structure Mapping Techniques.*, 2003.
- [Fred 04] Charles Fred, Martin Olivier, Cavazza Marc, Nandi Alok & Mari-chal Xavier. *Compelling Experiences in Mixed Reality Interactive Storytelling.* ??, 2004.
- [Frintrop 04] Simone Frintrop, Andreas Nüchter, Hartmut Surmann & Joachim Hertzberg. *Saliency-based Object Recognition in 3D Data*, 2004.
- [Fritz] G. Fritz, Ch. Seifert, L. Paletta & Horst Bischof. *Attentive Object Detection Using an Information Theoretic Saliency Measure.*
- [Fritz 04] Gerald Fritz, Christin Seifert, Lucas Paletta & Horst Bischof. *Entropy based saliency maps for object recognition.* ECOVISION, 2004.
- [Gaborski 03] Roger S. Gaborski, Vishal S. Vaingankar & Roxanne Canosa. *Goal Directed Visual Search Based on Color Cues : Co-operative Effects of Top-Down & Bottom-Up Visual Attention.* In Proceedings of the Artificial Neural Networks in Engineering, Rolla, Missouri, 2003.
- [Gao 04] Dashan Gao & Nuno Vasconcelos. *Discriminant Saliency for Visual Recognition from Cluttered Scenes.* In Proceedings of Neural Information Processing Systems (NIPS), 2004.
- [Garbay 00] C Garbay. Les systemes de vision, traité ic2, chapitre Architectures Logicielles et Contrôle dans les Systèmes de Vision, pages 197–249. Hermès, 2000.
- [Gibson 79] J.J. Gibson. *The ecological approach to visual perception.* Houghton Mifflin, 1979.
- [Gool 99] Luc Van Gool, Tinne Tuytelaars & Marc Pollefeys. *Adventurous tourism for couch potatoes.* In 8th International Conference on Computer Analysis of Images and Patterns, pages 98–107, 1999.
- [Gordon 89] I. E. Gordon. Theories of visual perception, chapitre The Gestalt theory, pages 46–75. New York : Wiley and Sons, 1989.
- [Graf 02] Birgit Graf & Oliver Barth. *Entertainment Robotics : Examples, Key Technologies and Perspectives*, 2002.
- [Grafe 02] M. Grafe, R. Wortmann & H. Westphal. *AR-based interactive exploration of a museum exhibit.* Augmented Reality Toolkit, The First IEEE International Workshop, 29 September 2002.
- [Grosz 86] Barbara J. Grosz & Candace L. Sidner. *Attention, Intentions, and the Structure of Discourse.* Computational Linguistics, vol. 12, no. 3, pages 175–204, 1986.

- [Grosz 95] Barbara J. Grosz, Scott Weinstein & Aravind K. Joshi. *Centering : A Framework for Modeling the Local Coherence of Discourse*. Computational linguistics, vol. 21, no. 2, 1995.
- [Halelamien 04] Neil S. Halelamien. *Visual routines for spatial cognition on a mobile robot*. Senior thesis, 2004.
- [Hanson 78] A. R. Hanson & E. M. Riseman. *VISIONS : a computer system for interpreting scenes*. Computer Vision Systems, pages 303–333, 1978. Academic Press.
- [Harrison 02] J. Harrison. *Approaches to Perception and Cognition. Philosophical, Early, Classic and Modern BC 600 to Today*. <http://www.cs.ubc.ca/~harrison/P202/websites.html>, 2002. PSYC202-005 : Perception and Cognition.
- [Hayhoe 02] Mary M. Hayhoe, Dana H. Ballard, Hiroyuki Shinoda Jochen J. Triesch, Pilar Aivar & Brain T. Sullivan. *Vision in Natural and Virtual Environments*. Eye Tracking Research and Applications Symposium, 2002.
- [Heinecke 95] Andreas M. Heinecke. *Evaluation of Hypermedia Systems in Museums*. Archive & Museum Informatics, pages 67–79, 1995.
- [Henderson 03a] J. Henderson & A. Hollingworth. *Eye movements and visual memory : detecting changes to saccade targets in scenes*, 2003.
- [Henderson 03b] John M. Henderson. *Human Gaze control during real-world scene perception*. TRENDS in Cognitive Sciences, vol. 7, no. 11, pages 498–504, 2003.
- [Herpers 98] R. Herpers & G. Sommer. *An attentive processing strategy for the analysis of facial features*. Face Recognition : From Theory to Applications, H. Wechsler et al. NATO ASI Series F, Springer-Verlag, pages 457–468, 1998.
- [Herviou 06] David Herviou. *La perception visuelle des entités autonomes en réalité virtuelle : Application à la simulation de trafic routier*. PhD thesis, Université de Bretagne Occidentale, December 2006.
- [Herwin van Welbergen 06] Dennis Reidsma Herwin van Welbergen Anton Nijholt & Job Zwiers. *Presenting in virtual worlds : An Architecture for a 3D Anthropomorphic Presenter*. IEEE Computer Society, SEPTEMBER/OCTOBER 2006.
- [Herzog 94] Gerd Herzog & Peter Wazinski. *Visual TRANslator : Linking Perceptions and Natural Language Descriptions*. Artificial Intelligence Review, vol. 8, no. 2/3, pages 175–187, 1994.
- [Hewett 01] Micheal Hewett. *Computational Perceptual Attention*, 2001.

- [Hill R.W. 99] Jr. Hill R.W. *Perceptual grouping and attention in a multi-agent world*. In Proceedings of the third annual conference on Autonomous Agents, pages 418–419, 1999.
- [Hollerer 99a] T. Hollerer, S. Feiner, T. Terauchi, G. Rashid & D. Hallaway. *Exploring MARS : Developing indoor and outdoor user interfaces to a mobile augmented reality system*, 1999.
- [Horswill 95] Ian Horswill. *Visual Routines and Visual Search : A Real-Time Implementation and an Automata-Theoretic Analysis*. In IJCAI, pages 56–63, 1995.
- [Horswill 97a] Ian Horswill. *Real-time Control of Attention and Behavior in a Logical Framework*. In W. Lewis Johnson & Barbara Hayes-Roth, éditeurs, Proceedings of the First International Conference on Autonomous Agents (Agents'97), pages 130–137, New York, 5–8, 1997. ACM Press.
- [Horswill 97b] Ian Horswill. *Visual architecture and cognitive architecture*. JE-TAI, vol. 9, no. 2-3, pages 277–292, 1997.
- [Hoult 04] Christopher Hoult. *Emotion in Speech Synthesis*, 6 May 2004.
- [Hovy] Eduard H. Hovy. *Planning Coherent Multisentential Text*.
- [Hovy 93] Eduard H. Hovy. *Automated Discourse Generation using Discourse Structure Relations*. Artificial Intelligence–Special Issue on Natural Language Processing, vol. 63, 1993.
- [Howarth 92] R. Howarth & H. Buxton. *An analogical representation of space and time*. Image and Vision Computing, vol. 10, no. 7, 1992.
- [Hughes 04] CE. Hughes, E. SMith, C. Stapleton & DE. Hughes. *Augmenting museum experiences with mixed reality*. In Proceedings of the KSCE 2004, 2004.
- [Ibanez 03] J. Ibanez, R. Aylett & R. Ruiz-Rodarte. *Storytelling in virtual environments from a virtual guide perspective*. Virtual Reality, vol. 7, pages 30–42, 2003.
- [Ibanez 04a] J. Ibanez & R.S. Aylett. *Welcome to my virtual place*. In L Sheremetov & M Alvarado, éditeurs, Proceedings of Workshops on Intelligent Computing, MICAI 2004, pages 316–323, Instituto Mexicano del Petroleo & Sociedad Mexicana de Intelligencia Artificial, AC., 2004.
- [Ibanez 04b] J. Ibanez, C. Delgado, R.S. Aylett & R. Ruiz. *Don't you escape. i'll tell you my story*. In Proceedings of MICAI, pages 49–58, Mexico City, 2004.

- [Itti 98] L. Itti, C. Koch & E. Niebur. *A model of saliency-based visual attention for rapid scene analysis*. In IEEE Transactions on Pattern Analysis and Machine Intelligence **20**, pages 1254–1259, November 1998.
- [Itti 99a] L. Itti & C. Koch. *A Comparison of Feature Combination Strategies for Saliency-Based Visual Attention Systems*. In SPIE Human Vision and Electronic Imaging (HVEI'99), San José, CA, pages 373–382, January 1999.
- [Itti 00b] L. Itti & C. Koch. *A saliency based search mechanism for overt and covert shifts of visual attention*. In Vision Research **40**, pages 1489–1506, May 2000.
- [Itti 05] L. Itti. *Models of Bottom-Up Attention and Saliency*. In L. Itti, G. Rees & J. K. Tsotsos, éditeurs, Neurobiology of Attention, pages 576–582. Elsevier, San Diego, CA, Jan 2005.
- [James 90] W. James. *The Principles of Psychology*, 1890.
- [Jensen 02] B. Jensen, G. Froidevaux, X. Greppin, A. Lorotte, L. Mayor, M. Meisser, G. Ramel & R. Siegwart. *The interactive autonomous mobile system RoboX*. In International Conference on Intelligent Robots and Systems, volume 2, pages 1221–1227, Lausanne, Switzerland, October 2002.
- [Joachims 97] Thorsten Joachims, Dayne Freitag & Tom M. Mitchell. *Web Watcher : A Tour Guide for the World Wide Web*. In IJCAI (1), pages 770–777, 1997.
- [Johnson 94] Michael Patrick Johnson, Pattie Maes & Trevor Darrell. *Evolving Visual Routines*. In Rodney A. Brooks & Pattie Maes, éditeurs, ARTIFICIAL LIFE IV, Proceedings of the fourth International Workshop on the Synthesis and Simulation of Living Systems, pages 198–209, MIT, Cambridge, MA, USA, 6-8 1994. MIT Press.
- [Johnson 96] Michael Patrick Johnson. *Automated creation of visual routines using genetic programming*. In 13th International Conference on Pattern Recognition (ICPR'96), volume 1, page 951, 1996.
- [Johnson 00] W. Lewes Johnson, Jeff W. Rickel & James C. Lester. *Animated Pedagogical Agent : Face-to-Face Interaction in Interactive Learning Environments*. International Journal of Artificial Intelligence in Education, vol. 11, pages 47–78, 2000.
- [Jonker 99] Catholijn M. Jonker, Jan Treur & Wouter C.A. Wijngaards. *An executable Model of the Interaction between Verbal and Non-Verbal Communication*. In Agent Communication Languages Workshop (ACL'99), 1999.

- [Jähne 99] B. Jähne, H. Haußecker & P. Geißler. Handbook of computer vision and applications. Academic Press, 1999. Three Volume set.
- [Kadobayashi 98] Rieko Kadobayashi & Kenji Mase. *Seamless Guidance by Personal Agent in Virtual Space Based on User Interaction in Real World*. In PAAM'98, London, UK, 23–25 March 1998.
- [KaewTraKulPong 01] P. KaewTraKulPong & R. Bowden. *An Improved Adaptive Background Mixture Model for Real-Time Tracking with Shadow Detection*. In 2nd European Workshop on Advanced Video Based Surveillance Systems, 2001.
- [Kahneman 92] Daniel Kahneman, Anne Treisman & Brian J. Gibbs. *The reviewing of object files : Object-specific integration of information*. Cognitive Psychology, vol. 24, no. 2, pages 175–219, April 1992.
- [Kelly 02] Lynda Kelly. *What do people want from their museum experience ?* Australian Museum Audience Research Centre, June 2002.
- [Kim 04a] Gunhee Kim, Woojin Chung, Kyung-Rock Kim, Munsang Kim, Sangmok Han & Richard H. Shim. *The Autonomous Tour-Guide Robot Jinny*. In Proceedings of 2004 IEEWRSJ International Conference on Intelligent Robots and Systems, 2004.
- [Klesen 03] Martin Klesen, Michael Kipp, Patrick Gebhard & Thomas Rist. *Staging exhibitions : methods and tools for modeling narrative structure to produce interactive performances with virtual actors*. Virtual Reality, 2003.
- [Koch 85] C. Koch & S. Ullman. *Shifts in Selective Visual Attention : Towards the Underlying Neural Circuitry*. Human Neurobiology, vol. 4, pages 219–227, January 1985.
- [Koda 96] T. Koda & P. Maes. *Agent with faces : The effect of personification*. In Proceedings of the Fifth IEEE International Workshop on Robot and Human Communication, Tsukuba, Japan, November 1996.
- [Koper 04] Rob Koper & Jocelyn Manderveld. *Educational modelling language : modeling reusable, interoperable, rich and personalised units of learning*. British Journal of Educational Technology, vol. 35, no. 5, pages 537–551, September 2004.
- [Kopp 05] S. Kopp, L. Gesellensetter, N. Krämer & I. Wachsmuth. *A conversational agent as museum guide – design and evaluation of a real-world application*. Panayiotopoulos et al. (eds.) : Intelligent Virtual Agents, LNAI 3661, pp. 329–343, Berlin : Springer-Verlag, 2005.

- [Kuno 07] Yoshinori Kuno, Kazuhisa Sadazuka, Michie Kawashima, Keiichi Yamazaki, Akiko Yamazaki & Hideaki Kusuoka. *Museum Guide Robot Based on Socological Interaction Analysis*. In CHI 2007, pages 1191–1194, San Jose – California, USA, April 28 – May 3 2007. ACM.
- [Lemmetty 99] Sami Lemmetty. Review of speech synthesis technology. Master's thesis, Helsinki University of Technology, Departement of Electrical and Communication Engineering, 1999.
- [Lepouras 01] George Lepouras, Dimitrios Charitos, Costas Vassilakis, Anna Charissi & Leda Halatsi. *Building a VR-Museum in a Museum*. In VRIC, Virtual Reality International Conference, Laval Virtual, Laval, May 16–18 2001.
- [Leslie 98] Alam M. Leslie, Fei Xu, Patrice D. Tremoulet & Brian J. Scholl. *Indexing and the Object Concept : Developping "What" and "Where" systems*. Trends in Cognitive Science, vol. 2, no. 1, pages 10–18, January 1998.
- [List 04] Thor List & Robert B. Fisher. *CVML An XML-based Computer Vision Markup Language*. In Proceedings of the Pattern Recognition, 17th International Conference, 2004.
- [List 05] Thor List, José Bins, Robert B. Fisher, David Tweed & Kristinn R. Thórisson. *Two Approaches to a Plug-and-Play Vision Architecture – Caviar And Psyclone*. AAAI-05 Workshop on Modular Construction of Human-Like Intelligence, July 2005.
- [Louchart 04] Sandy Louchart & Ruth Aylett. *Narrative theory and emergent interactive narrative*. International Journal of Continuing Engineering Education and Lifelong Learning, vol. 14, no. 6, pages 506–518, 2004.
- [Lu 00] Shijian Lu, François Paradis, Cécile Paris, Stephen Wan, Ross Wilkinson & MingFang Wu. *Generating Personal Travel Guides from Discourse Plans*. In Adaptive hypermedia and adaptive web-based systems., volume 1892, pages 392–395, Trento , ITALIE, August 20-30 2000.
- [MacIntyre 96] Blair MacIntyre & Steven Feiner. *Future Multimedia User Interfaces*. Multimedia Systems, vol. 4, pages 250–268, 1996.
- [Maes 94] Pattie Maes. *Agents that reduce work and information overload*. Communications of the ACM, vol. 37, no. 7, 1994.
- [Maes 95] Pattie Maes. *Artificial Life Meets Entertainment : Lifelike Autonomous Agent*. Communications of the ACM, vol. 58, no. 11, pages 108–114, November 1995.

- [Malaka 00] Rainer Malaka & Alexander Zipf. *Deep Map - Challenging IT research in the framework of a tourist information system*. In Proceedings of ENTER 2000, 7th. International Congress on Tourism and Communications Technologies in Tourism, 2000.
- [Mann 87] William C. Mann & Sandra A. Thompson. Rhetorical structure theory : A theory of text organization. 1987.
- [Marchand 96] Eric Marchand. *Stratégies de perception par vision active pour la reconstruction et l'exploration de scènes statiques*. PhD thesis, Université de Rennes 1, 1996.
- [Marr 82] D. Marr. Vision : A computational investigation into the human representation and processing of visual information. W. H. Freeman, 1982.
- [Marti 99] P. Marti, A. Rizzo, L. Petroni, G. Tozzi & M. Diligenti. *Adapting the Museum : A Non-Intrusive User Modeling Approach*, 1999.
- [Marucci 00b] Luisa Marucci & Fabio Paternò. *Designing an Adaptative Virtual Guide for Web Applications*. In 6th ERCIM Workshop "User Interfaces for All", Florence, Italy, 25-26 October 2000.
- [Marucci 00c] Luisa Marucci & Fabrizio Paternò. *Designing an Agent to Support Interactive Access to Museum Information*. WOA, 2000.
- [Mateas 98] Michael Mateas & Phoebe Sengers. *Narrative Intelligence*. AAAI, 1998.
- [Matsuyama 89] T. Matsuyama. *Expert systems for image processing : knowledge-based composition of image analysis processes*. Computer Vision, Graphics and Image Processing, vol. 48, no. 1, 1989.
- [Mccallum 94] Andrew Mccallum. *Short-Term Memory in Visual Routines for Off-Road Car Chasing*, 1994.
- [McKeown 82] Kathleen R. McKeown. *The TEXT System for Natural Language Generation : An Overview*. In Proceedings of the 20th Annual Meeting of the ACL, Toronto, June 1982. Association of Computational Linguistics.
- [McKeown 85] Kathleen R. McKeown. Text generation : Using discourse strategies and focus constraints to generate natural language texts. Cambridge University Press, 1985.
- [Mellish 98] Chris Mellish, Mick O'Donnell, Jon Oberlander & Alistair Knott. *An architecture for Opportunistic Text Generation*. In Workshop on Natural Language generation, 1998.
- [Milanese 94] R. Milanese, H. Wechsler, S. Gil, J-M. Bost & T. Pun. *Integration of Bottom-Up and Top-Down Cues for Visual Attention Using Non-Linear Relaxation*. IEEE Conf. on Computer Vision and Pattern Recognition, 1994.

- [Milgram 94] Paul Milgram & Fumio Kishino. *A taxonomy of Mixed Reality Visual Displays*. IEICE Transactions on Information Systems, vol. E77-D, no. 12, pages 1321–1329, December 1994.
- [Miller 56] George A. Miller. *The magical number seven, plus or minus two : Some limits on our capacity for processing information*. Psychological Review, vol. 63, pages 81–97, 1956.
- [Moisan 95] Sabine Moisan, Régis Vincent, Monique Thonnat, Véronique Clément & John van den Elst. *Manuel de référence du logiciel OCAPI Version 2.0*. Rapport technique 0183, Institut National de Recherche en Informatique et en Automatique, 1995.
- [Moraes 99] A.C. da Rocha Costa-A.C. Moraes M.C. Bertoletti. *Virtual guides to assist visitors in the SAGRES virtual museum*. In Computer Science Society, 1999. Proceedings. SCCC '99. XIX International Conference of the Chilean, pages 35–42, Talca, Chile, 1999.
- [Nagao 95] K. Nagao & J. Rekimoto. *Ubiquitous talker : Spoken Language interaction with real world objects*. In IJCAI-95, pages 1284–1290, 1995.
- [Navalpakkam 02] V. Navalpakkam & L. Itti. *A goal oriented attention guidance model*. In Proc. 2nd Workshop on Biologically Motivated Computer Vision (BMCV'02), pages 453–461, Tuebingen, Germany, November 2002.
- [Navalpakkam 05] V. Navalpakkam & L. Itti. *Modeling the Influence of Task on Attention*. Vision Research, vol. 45, pages 205–231, 2005.
- [Newcombe 04] Richard A. Newcombe. *A Biologically Inspired Visual Attention Framework for Computer Vision*. In partial fulfilment of the requirements for the degree of BSc(Hons) G5G8 Robotics and Intelligent Machines, March 2004.
- [Nijholt 03b] Anton Nijholt. *Disappearing Computers, Social Actors and Embodied Agents*. In T.L. Kunii, S. Hock Soon & A. Sourin, éditeurs, International Conference on CYBERWORLDS, numéro ISBN 0-7695-1922-9, pages 128–134, Singapore, December 2003. IEEE Computer Society Press Los Alamitos USA.
- [Noma 00] Tsukasa Noma, Liwei Zhao & Norman I. Badler. *Design of a Virtual Human Presenter*. IEEE Computer Graphics and Applications, vol. 20, no. 4, pages 79–85, J1/Aug 2000.
- [Nourbakhsh 99] Illah R. Nourbakhsh, Judith Bobenage, Sebastien Grange, Ron Lutz, Roland Meyer & Alvaro Soto. *An affective mobile robot educator with a full-time job*. <http://www.cs.cmu.edu/~illah/SA-GE/>, 1999.

- [Nwana 96] Hyacinth S. Nwana. *Software Agents : An Overview*. Knowledge Engineering Review, vol. 11, no. 3, pages 1–40, September 1996. Cambridge University Press.
- [Oksama 04] L. Oksama & J. Hyöna. *Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition ? An individual difference Approach*. Visual Cognition, vol. 11, pages 631–671, 2004.
- [Olivia 03] Aude Olivia, Antonio Torralba, Monica S. Castelhana & John M. Henderson. *Top-Down Control of Visual Attention in Object Detection*. In Proceedings of the IEEE International Conference on Image Processing, volume 1, pages 253–256, Barcelona, Spain, September 2003.
- [Ouerhani 03] Nabil Ouerhani. *Visual Attention : From Bio-Inspired Modeling to Real-Time Implementation*. PhD thesis, Institut de Microtechnique Université de Neuchatel, 2003.
- [Padgham 97] Lin Padgham & Guy Taylor. *A system for modelling agents having emotions and personality*, 1997.
- [Paletta 05] Lucas Paletta, Erich Rome & Hilary Buxton. *Attention Architectures for Machine Vision and Mobile Robots*. Neurobiology of Visual Attention, 2005.
- [Palmer 99] Stephen E. Palmer. *Vision science : Photons to phenomenology*. MIT Press, 1999.
- [Panayiotopoulos 99] T. Panayiotopoulos, N. Zacharis & S. Vosinakis. *Intelligent Guidance in a Virtual University*. In S. Tzafestas, editeur, *Advances in Intelligent Systems - Concepts, Tools and Applications*, pages 33–42. Kluwer Academic Press, 1999.
- [Pape 98] Dave Pape, Tomoko Imai, Josephine Anstey, Maria Roussou & Tom DeFanti. *XP : An Authoring System for Immersive Art Exhibitions*. In Fourth International Conference on Virtual Systems and Multimedia, Gifu, Japan, 1998.
- [Paris 01a] Paris, Wan, Wilkinson & M-F Wu I. *Generating Personal Travel Guides ? on demand, on the fly, on the go*. In Proceedings of the Australian Workshop on Natural Language Processing 2001, Sydney, Australia, 2001.
- [Paris 01b] Paris, Wan, Wilkinson & M-F Wu. *Generating Personal Travel Guides ? and who wants them ?* In Proceedings of the International Conference on User Modelling (UM2001), pages 13–18, Sonthofen, Germany, July 2001.
- [Paulus 99] D. Paulus, U. Ahlrichs, B. Heigl, J. Denzer, J. Hornegger & H. Niemann. *Active Knowledge-based scene analysis*. In H.E.

- Christensen, éditeur, Proceedings of the first international conference on computer vision systems, ICVS-99, pages 180–199, 1999.
- [Pelachaud 04] Catherine Pelachaud, Annelies Brafford, Gaspard Breton, Nicolas Ech Chafai, Sylvie Gibet, Jean-Claude Martin, Sébastien Maubert, Magalie Ochs, Danielle Pelé, Alexandre Perrin, Matthieu Raynal, Lionel Réveret & David Sadek. *Agent Conversationnels : Systèmes d'animation, Modélisation des comportements multimodaux, Applications : Agents pédagogiques et Agents signeurs*. Rapport technique, AS Humain Virtuel, thème 4 : Agent conversationnels, 2004.
- [Pernin 04] J-P. Pernin & A Lejeune. *Dispositifs d'apprentissage instrumentés par les technologies : vers une ingénierie centrée sur les scénarios*. In Actes du colloque TICE, 2004.
- [Paternier 07] A. Paternier, X. Righetti, M. Hopmann, D. Thalmann, M. Repetto, G. Papagiannakis, P. Davy, M. Lim, N. Magnenat-Thalmann, P. Barsocchi, T. Fragopoulos, D. Serpanos, Y. Gialelis & A. Kirykou. *Chloe@University : an indoor, mobile mixed reality guidance system*. In Proceedings of the 2007 ACM symposium on Virtual Reality, 2007.
- [Peters 03] Christopher Peters & Carol O'Sullivan. *Bottom-Up Visual Attention for Virtual Human Animation*. In 16th International Conference on Computer Animation and Social Agents, pages 111–117, May 2003.
- [Picot 06] A. Picot, G. Bailly, F. Elisei & S. Raidt. *Scrutation de scènes naturelles par un agent conversationnel animé*. Toulouse, France, 2006.
- [Pietsch 04] Kay Pietsch & Katrin Steinmann. Potentials of virtual museums - media-specific conception of cultural learning environments. Master's thesis, University of Applied Sciences, Kiel, Germany, August, 19 2004.
- [Pisan 95] Yusuf Pisan. *A visual routines based model of graph understanding*. In Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society, pages 692–697, Pittsburgh, 1995.
- [Pollefeys 01] Marc Pollefeys, Luc Van Gool, Ive Akkermans, Dirk De Becker & Kris Demuynck. *A Guided Tour to Virtual Sagalassos*, 2001.
- [Pomerantz 89] James R. Pomerantz & Edward A. Pristach. *Emergent Features, Attention, and Perceptual Glue in Visual Form Perception*. Journal of Experimental Psychology : Human Perception and Performance, vol. 15, no. 4, pages 635–649, 1989.
- [Popovici 03] Dorin-Mircea Popovici, Serge Morvan, Eric Maisel & Jacques Tisseau. *Interactive Distributed Guided Tours of Historical Sites*. In

- Proceedings of the 2003 International Conference on Cyberworlds (CW'03), 2003.
- [Pospischil 02] Günther Pospischil, Martina Umlauf & Elke Michlmayr. *Designing LoL@, a Mobile Tourist Guide for UMTS*. In Proceedings of the 4th International Symposium on Mobile Human-Computer Interaction, 2002.
- [Pylyshyn 89] Zenon Pylyshyn. *The Role of Location Indexes in Spatial Perception : A Sketch of the FINST Spatial Index Model*. Cognition, vol. 32, no. 1, pages 65–97, 1989.
- [Pylyshyn 94] Z. Pylyshyn, J. Burkell, B. Fisher, C. Sears, W. Schmidt & L. Trick. *Multiple parallel access in visual attention*. Canadian Journal of Experimental Psychology, vol. 48, no. 2, pages 260–283, 1994.
- [Ramström 04a] Ola Ramström. *Visual Attention using Game Theory*. Licentiate thesis, KTH, Numerical Analysis and Computer Science, Stockholm, Sweden, September 2004.
- [Rao 88] A. R. Rao & R. Jain. *Knowledge representation and control in computer vision systems*. IEEE Expert, vol. 3, no. 1, pages 64–79, 1988.
- [Rao 98] Satyajit Rao. *Visual Routines and Attention*. PhD thesis, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, 1998.
- [Raptis 05] Dimitrios Raptis, Nikolaos Tselios & Nikolaos Avouris. *Context-based Design of Mobile Applications for Museums : A Survey of Existing Practices*. In MobileHCI'05, Salzburg, Austria, September, 19–22 2005. ACM.
- [Rayward 99] M.B. Rayward W.B. and Twidale. *From Docent to Cyberdocent : Education and Guidance in the Virtual Museum*. Archives and Museum Informatics, Volume 13, Number 1, 1999 , pp. 23-53(31), 1999.
- [Redeker 00] G. Redeker. Abduction, belief and context in dialogue. studies in computational pragmatics, chapitre Coherence and structure in text and discourse., pages 233–263. Amsterdam : Benjamins, 2000.
- [Reece 95] D. A. Reece & S. A. Shafer. *Control of Perceptual Attention in Robot Driving*. Artificial Intelligence, vol. 78, pages 397–430, 1995.
- [Reeves 96] Byron Reeves & Clifford Nass. The media equation : How people treat computers, television, and new media like real people and places. CLSI Lecture notes, 1996.

- [Rensink 97] Rensink, O'Regan & Clark. *To See or Not to See : The Need for Attention to Perceive Changes in Scenes*. In Psychological Science, volume 8, pages 368–373. 1997.
- [Rensink 00] R.A. Rensink. *The Dynamic Representation of Scenes*. Visual Cognition, vol. 7, pages 17–42, 2000.
- [Reynolds 87] Craig W. Reynolds. *Flocks, Herds, and Schools : A Distributed Behavioral Model*. Computer Graphics, vol. 21, no. 4, pages 25–34, 1987.
- [Rickel 98] Jeff Rickel & W. Lewis Johnson. *STEVE : A Pedagogical Agent for Virtual Reality*. In Katia P. Sycara & Michael Wooldridge, éditeurs, Proceedings of the 2nd International Conference on Autonomous Agents (Agents'98), pages 332–333, New York, 9–13, 1998. ACM Press.
- [Roelfsema 03] Paul S. ; Spekreijse-Henk Roelfsema Pieter R. ; Khayat. *Subtask sequencing in the primary visual cortex*. In Proceedings of the National Academy of Science, volume 100, pages 5467–5472, 2003.
- [Rosenfeld 99] A. Rosenfeld. *Image Analysis and Computer Vision*. Computer Vision and Image Understanding, vol. 78, pages 222–302, 1999.
- [Rothenstein 06] Albert L. Rothenstein & John K. Tsotsos. *Attention links sensing to recognition*. Image and Vision Computing, pages 1–13, 2006.
- [Rullen 00] R. Van Rullen & C. Koch. Visual attention and visual awareness, chapitre 3. 2000.
- [Salgian 98a] Garbis Salgian. *Tactical Driving Using Visual Routines*. PhD thesis, University of Rochester, Rochester, New York, 1998.
- [Salgian 98b] Garbis Salgian & Dana H. Ballard. *Visual Routines for autonomous driving*. In 6-th ICCV, pages 876–882, Bombay, India, 1998.
- [Salomon 89] G. Salomon, T. Oren & K. Kreitman. *Using guides to explore multimedia databases*. In Proceedings of the Twenty-Second Annual Hawaii International Conference on System Sciences, 1989.
- [Sandalky 95] Fadi Sandalky. *Contribution à la mise en oeuvre d'une architecture à base de connaissances pour l'interprétation de scènes 2D et 3D*. PhD thesis, Université de Nice, Nice, France, 1995.
- [Schiele 01] B. Schiele, T. Jebara & N. Oliver. *Sensory-augmented computing : wearing the museum's guide*. IEEE Micro, vol. 21, no. 3, pages 44–52, 2001.
- [Schmalstieg 05] Dieter Schmalstieg & Daniel Wagner. *A Handheld Augmented Reality Museum Guide*, 2005.

- [Schröder 01] M. Schröder. *Emotional speech synthesis : a review*. In Proceedings of Eurospeech, 2001.
- [Shimon 92] E. Shimon. *Visual perception*. In Stuart C. Shapiro, editeur, Encyclopedia of Artificial Intelligence, chapitre Encyclopedia of Artificial Intelligence, pages 1655–1664. Wiley-Interscience, New-York, 1992.
- [Soyer 00] C. Soyer, H. I. Bozma & Y. I Stefanopoulos. *A new memory model for selective perception systems*. In Proceedings of the 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2000.
- [Sperling 60] Georges Sperling. *The information available in brief presentation*. Psychological Monographs, vol. 74, no. 11, pages 1–30, 1960.
- [Sprague 01] Nathan Sprague & Dana Ballard. *A Visual Control Architecture for a Virtual Humanoid*. In Proceedings of the IEEE-RAS International Conference on Humanoid Robots, 2001.
- [Starner 97] Thad Starner, Steve Mann, Bradley Rhodes, Jeffrey Levine, Jennifer Healey, Dana Kirsch, Roz Picard & Alex Pentland. *Augmented Reality Through Wearable Computing*. In Presence : Teleoperators and Virtual Environments, 1997.
- [Stentiford 01] F. W. M. Stentiford. *An Evolutionary Programming Approach to the Simulation of Visual Attention*. In Proceedings of the 2001 Congress on Evolutionary Computation CEC2001, pages 851–858, COEX, World Trade Center, 159 Samseong-dong, Gangnam-gu, Seoul, Korea, 27-30 2001. IEEE Press.
- [Stock 04] O. Stock & M. Zancanaro. *Intelligent interactive information presentation for cultural tourism*. In International Workshop on Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems, Copenhagen, June 2004. Invited talk.
- [Stone 96] Brian A. Stone & James C. Lester. *Dynamically sequencing an animated pedagogical agent*, 1996.
- [Sumi 98] Yasuyuki Sumi, Tameyuki Etani, Sidney Fels, Nicolas Simonet, Kaoru Kobayashi & Kenji Mase. *C-MAP : Building a Context-Aware Mobile Assistant for Exhibition Tours*. Lecture Notes In Computer Science, vol. 1519, pages 137–154, 1998.
- [Swain 93] Michael J. Swain & Markus A. Stricker. *Promising directions in active vision*. International Journal of Computer Vision, vol. 11, no. 2, pages 109–126, October 1993.
- [Tanaka-Ishii 98] Kumiko Tanaka-Ishii, Koiti Hasida & Itsuki Noda. *Reactive Content Selection in the Generation of Real-time Soccer Commentary*. In COLING-ACL, pages 1282–1288, 1998.

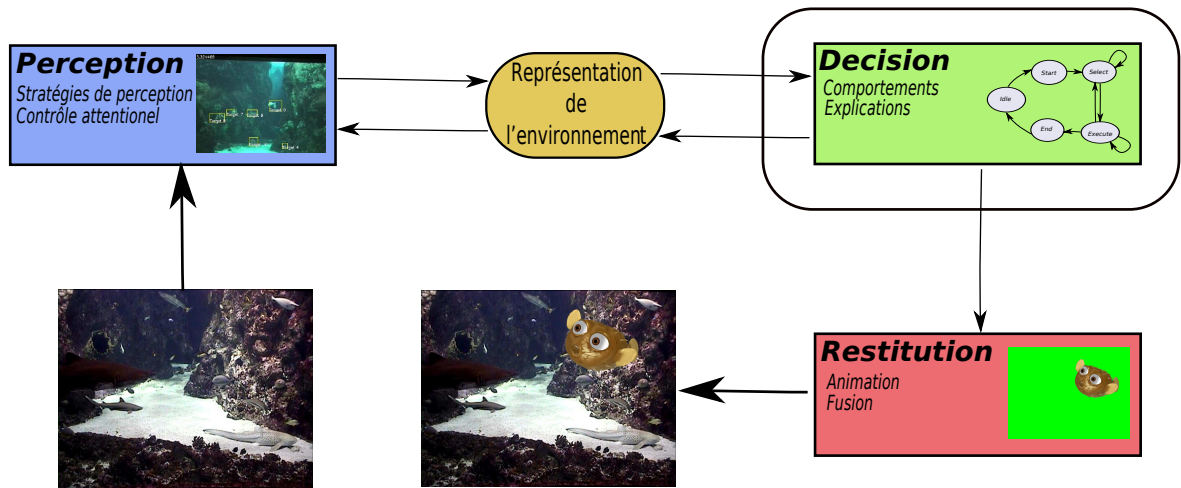
- [Tanaka-Ishii 00] Kumiko Tanaka-Ishii & Eibe Ian Frank. *Multi-agent explanation strategies in real-time domains*. In Proceedings of the 38th Annual Meeting on Association for Computational Linguistics, pages 158–165, 2000.
- [Tchounikine 02] Pierre Tchounikine. *Quelques éléments sur la conception et l'ingénierie des EIAH*. In Actes du GDR I3, pages 233–245, 2002.
- [Theune 01] M. Theune. *ANGELICA : choice of output modality in an embodied agent*. In International Workshop on Information Presentation and Natural Multimodal Dialogue (IPNMD-2001), pages 89–93, Verona, Italy, December 2001.
- [Theune 05] Mariët Theune, Dirk Heylen & Anton Nijholt. *Generating embodied information presentations*. O. Stock and M. Zancanaro (eds.), Multimodal Intelligent Information Presentation, Kluwer Academic Publishers, ISBN 1-4020-3049-5, pages 47-70., 2005.
- [Thomas 02] Bruce Thomas, Ben Close, John Donoghue, John Squires, Philip De Bondi & Wayne Piekarski. *First Person Indoor/Outdoor Augmented Reality Application : ARQuake*. Personal and Ubiquitous Computing, vol. 6, pages 75–86, 2002.
- [Thonnat 93] Monique Thonnat, Vincent Clément & John van den Elst. *Supervision of perception tasks for autonomous systems : the OCAPI approach*. Rapport de recherche 2000, INRIA, 1993.
- [Thrun 99] Sebastian Thrun, Maren Bennewitz, Wolfram Burgard, Armin B. Cremers, Frank Dellaert, Dieter Fox, Dirk Hahnel, Charles Rosenberg, Nicholas Roy, Jamieson Schultel & Dirk Schulz. *MINERVA : A Second-Generation Museum Tour-Guide Robot*. In Proceedings of the 1999 IEEE International Conference on Robotics & Automation, Detroit, Michigan, May 1999.
- [Tisseau 01] J. Tisseau. *Autonomie in virtuo*. Habilitation a Diriger des Recherches, December, 6 2001.
- [Torralba 03] A. Torralba. *Modeling global scene factors of Attention*. Journal of Optical Society of America, vol. 20, no. 7, 2003. JOSA-A.
- [Torralba 05] Antonio Torralba. Contextual influences on saliency, chapitre 96, pages 586–592. Academic Press / Elsevier, 2005.
- [Treisman 80] Anne M. Treisman & Garry Gelade. *A feature-integration theory of attention*. Cognitive Psychology, vol. 12, no. 1, pages 97–136, January 1980.
- [Treue 03] Stefan Treue. *Visual attention : the where, what, how and why of saliency*, 2003.
- [Tricot 99] André Tricot & Alain Rufino. *Modalités et scénarios d'interaction dans les hypermédias d'apprentissage*. Revue des sciences de l'éducation, vol. 25, no. 1, pages 105–129, 1999.

- [Triesch 03] J. Triesch, D. Ballard, M. M. Hayhoe & B. T. Sullivan. *What You See is What You Need*. Journal of Vision, vol. 3, no. 1, pages 86–94, 2003.
- [Tsapatsoulis 03] K. Rapantzikos N. Tsapatsoulis. *On the Implementation of Visual Attention Architectures*. In Tales of the Disappearing Computer,, Santorini, June 2003 2003.
- [Tsotsos 89] J. K. Tsotsos. *The complexity of perceptual search tasks*. In Proceedings of the Eleventh International Joint Conference on Artificial Intelligence, pages 1571–1577, Detroit, Michigan*, 1989.
- [Tsotsos 92] John K. Tsotsos. Encyclopedia of artificial intelligence 2nd edition, volume 1, chapitre Image Understanding, pages 641–652. Wiley-Interscience, 1992.
- [Tsotsos 01] John K. Tsotsos. Complexity, vision and attention, chapitre 6, pages 105–128. Cambridge University Press, 2001.
- [Ullman 84] Shimon Ullman. *Visual Routines*. Cognition, vol. 18, pages 97–159, 1984.
- [van Es 06] René van Es & Rob Koper. *Testing the pedagogical expressivness of IMS LD*. Educational Technology & Society, vol. 9, no. 1, pages 229–249, 2006.
- [van Riel 06] C. van Riel, A. Eliëns & Y. Wang. *Exploration and guidance in media-rich information spaces : the implementation and realization of guided tours in digital dossiers*. In International Conference on Multidisciplinary Information Sciences and Technologies, 2006.
- [Vernon 04] David Vernon. Cognitive vision - the development of a discipline. 2004.
- [Veyret 04a] Morgan Veyret. Focalisation de l’attention visuelle. Rapport bibliographique de DEA, Janvier 2004.
- [Veyret 04b] Morgan Veyret. Modèle de focalisation de l’attention visuelle. Master’s thesis, Université de Rennes 1 - ENIB, 2004.
- [Veyret 04c] Morgan Veyret & Eric Maisel. *Simulation de la focalisation de l’attention visuelle : application à la simulation d’automobilistes virtuels*. In AFIG, Poitiers, 2004.
- [Vlahakis 02] Vassilios Vlahakis, Nikolaos Ioannidis, John Karigiannis, Manolis Tsotros, Michael Gounaris, Didier Stricker, Tim Gleue, Patrick Daehne & Luís Almeida. *Archeoguide : An Augmented Reality Guide for Archaeological Sites*. Computer Graphics in Art History and Archaeology, vol. 2, pages 52–60, September/October 2002.

- [Wachsmuth 97] Ipke Wachsmuth, Britta Lenzmann, Tanja Jording, Bernhard Jung, Marc Latoschik & Martin Frohlich. *A Virtual Interface Agent and its Agency*. In Proceedings of the first international conference on Autonomous agents, pages 516 – 517, Marina del Rey, California, United States, 1997.
- [Walther 02] D. Walther, L. Itti, M. Riesenhuber, T. Poggio & C. Koch. *Attentional selection for object recognition - a gentle way*, 2002.
- [Walther 04a] D. Walther, U. Rutishauser, C. Koch & P. Perona. *On the usefulness of attention for object recognition*. In Workshop on Attention and Performance in Computational Vision, 2004.
- [Walther 05] Dirk Walther, Ueli Rutishauser, Christof Koch & Pietro Perona. *Selective visual attention enables learning and recognition of multiple objects in cluttered scenes*. Computer Vision and Image Understanding, vol. 100, pages 41–63, 2005.
- [Weiser 91] Mark Weiser. *The computer for the Twenty-First Century*. Scientific American, pages 94–10, September 1991.
- [Welch 01] Greg Welch & Gary Bishop. *An Introduction to the Kalman Filter*. SIGGRAPH 2001, Course 8, 2001.
- [Wheeler 94] M. Wheeler. *Active Perception in Meaningful Worlds*. Brighton, UK, School of Cognitive and Computing Sciences, University of Sussex, 1994.
- [Willeke 01] Thomas Willeke, Clay Kunz & Illah Nourbakhsh. *The History of the Mobot Museum Robot Series : An Evolutionary Study*. In Proceedings of FLAIRS, 2001.
- [Wolfe 94] J.M. Wolfe. *Guided Search 2.0 : A Revised Model of Visual Search*. Psychonomic Bulletin & Review, vol. 1, no. 2, pages 202–238, 1994.
- [Wolfe 97] Jeremy M. Wolfe & Sara C. Bennett. *Preattentive Object Files : Shapeless Bundles of Basic Features*. Vision, vol. 37, no. 1, pages 25–43, 1997.
- [Yarbus 67] A. L. Yarbus. *Eye movements and vision*. Plenum, 1967. Originally published in Russian 1962.
- [Yi 05] W. Yi & D. H. Ballard. *Routine Based Models of Anticipation in Natural Behaviors*. In AAAI Fall Symposium, From Reactive to Anticipatory Cognitive Embodied Systems, Arlington, VA, USA, November 2005.
- [Yuan 03] Xiang Yuan & Yam San Chee. *Embodied Tour Guide in an Interactive Virtual Art Gallery*. In T. L. Kunii, Seah, H.S. & Sourin A., éditeurs, International Conference on CyberWorlds 2003, pages 432–439, Singapore, Los Alamitos, 2003. IEEE Computer Society.

- [Zheng 05b] Lihong Zheng & Xiangjian He. *Classification Techniques in Pattern Recognition*. In WSCG'2005, Plzen, Czech Republic., 2005.

Définition de la visite guidée



A.1 Introduction

La mise en place d'une visite guidée repose sur la définition de l'ensemble des données de discours, du comportement du guide virtuel et d'experts capables de guider le processus de sélection des explications présentées au visiteur. Ces différentes définitions se font selon une syntaxe décrite dans la suite de cette annexe. La définition du discours s'articule principalement autour de la définition des éléments de discours puis la structuration de ces éléments entre eux au sein d'explications et de transitions.

A.2 Définition des éléments de discours

Lors de la conception du moteur de discours nous avons souhaité mettre l'accent sur l'aspect auteur. L'idée étant de proposer une approche offrant un contrôle complet par les auteurs à

la fois sur le contenu du discours et sur la présentation de ce contenu au visiteur.

La définition d'un élément de discours permet aux auteurs de décrire la façon dont une information spécifique doit être présentée au visiteur. Pour cela il peut spécifier un ou plusieurs scripts d'animation déclarant ainsi différentes alternatives pour la présentation d'informations.

Cette définition s'effectue de la façon présentée dans le listing A.1. La propriété la plus intéressante ici est la liste d'animations. Dans l'exemple proposé, celle-ci ne définit qu'un unique script que nous allons détailler. Un script d'animation est un ensemble de paires (*condition, action*) définissant le déclenchement de commandes destinées au serveur d'animation. Ici, le script présenté déclenche la lecture d'un document sonore sans condition (donc dès le début de l'exécution du script d'animation) puis 3,5 secondes après le début de cette exécution, l'animation « hello » est déclenchée. Le dernier événement devant s'exécuter à 7,2 secondes après le début de l'exécution du script permet de s'assurer que l'exécution du script est effective jusqu'à la fin de la lecture du fichier son déclenché à son commencement.

Les différentes animations envisageables dépendent des primitives multimédia et d'animation que propose le module de rendu utilisé pour le guide virtuel. La totalité des possibilités du module utilisé dans le cadre du projet SIRENE sont présentées dans l'annexe B.

```
(defelement
: id 'element-unique-identifier
: actions (list 'do-something 'do-something-else)
: animations (list
(animation-script
(animation-step
(t)
((play-sound "some-sound-file"))))
(animation-step
((at-time 3.5)); in seconds
((play-animation "hello")))
(animation-step
((at-time 7.2)); in seconds
((lambda () '())))) )); wait the end of the sound file playing
```

LISTING A.1 – Exemple de déclaration d'un élément de discours.

A.3 Sujets, explications et transitions

Le discours est basé sur l'organisation des éléments de discours définis précédemment au sein de sujets et de transitions.

A.3.1 Définition d'une explication

A.3.2 Définition d'un sujet

La définition d'un sujet et de l'explication qui lui est associé permet aux auteurs de décrire les relations entre éléments de discours et la façon dont ceux-ci s'organisent.

```
(defscenario
  :id 'scenario-unique-identifier
  :adjacency ((element-1 element-2 element-3)
              (element-2 element-3)
              (element-3 nil))
  :choice-funcs (#'random-nth #'random-nth))
```

LISTING A.2 – Exemple de déclaration d'une explication (scénario d'explication).

Cette définition s'effectue suivant la syntaxe présentée dans le listing A.2. Le graphe décrivant l'explication est défini par une liste d'adjacence qui référence les élément de discours concernés grâce à leur identifiant unique. Pour chaque noeud de ce graphe, une fonction de choix est spécifiée (ici un choix aléatoire parmi l'ensemble des successeurs).

La définition d'un sujet associé à ce scénario permet de déclarer un certain nombre de mots-clés utilisés lors de la sélection des informations à présenter au visiteur.

```
(defsubject
  :id 'subject-unique-identifier
  :scenarios ('scenario-unique-identifier)
  :topics '(topic-1 topic-2 topic-3))
```

LISTING A.3 – Exemple de déclaration d'un sujet.

Cette définition est faite de la façon présentée dans le listing A.3. Une liste permet de définir l'ensemble des explications associées au sujet en question. Les différents thèmes (mots-clés) permettent de définir des informations utilisées lors du choix de l'explication à donner au visiteur. La possibilité de définir plusieurs explications pour un même sujet permet aux auteurs d'envisager par exemple des explications adaptées à différents types de visiteurs (enfants ou adulte par exemple), néanmoins cette fonctionnalité n'est pour le moment pas mise en œuvre de façon concrète lors de l'utilisation de ces données par le guide virtuel.

A.3.3 Définition d'une transition

La passage d'un sujet à un autre durant la visite guidée repose sur l'utilisation de transitions que les auteurs peuvent définir suivant une syntaxe similaire.

Cette syntaxe est présentée dans le listing A.4. La déclaration d'une transition permet de définir une liste d'éléments de discours équivalents pouvant être utilisés lors de l'exécution de la transition et les conditions d'utilisation de la dite transition (ici lors de la disparition d'une entité visuelle en cours d'explication).

```
(deftransition
: id 'transition-unique-identifier
: elts '(discourse-element-10 discourse-element-24)
: use-fun (entity-disappeared-p))
```

LISTING A.4 – Exemple de déclaration d'une transition.

A.3.4 Prototypage

Les données du discours sont créées de façon simplifiée par les personnes du service pédagogique d'Océanopolis à l'aide d'un fichier de configuration permettant de générer la déclaration des données réelles de la visite.

```
% Définition d'un sujet -----
#session-introduction
% Définition du graphe du scénario associé
((1 2 3)
 (2 3 4)
 (3 4)
 (4 nil))
% Définition des éléments de discours
-Salut! C'est moi Didon, le diodon. Je suis un poisson porc-épic
et j'habite ce lagon de l'océan Pacifique.
-Ici c'est mon domaine et je connais tout les poissons.
-Tu vois, je suis entouré de requins, des petits et des gros avec
de longues dents pointues...mais, qu'ils se méfient, moi aussi
j'ai des piquants pointus.
-Si tu veux, je t'emmène à la découverte de tout ce petit monde.

% Définition d'une transition -----
#transition-un-nom
% Définition des conditions d'utilisation
(entity-disappeared-p)
% Définition des éléments de discours
-Ho! Il s'en va, laissons le, on le reverra plus tard.
-Mais? Ou est-il passé? On le retrouvera sûrement plus tard.
```

LISTING A.5 – Syntase utilisée pour la définition des données de discours par les auteurs.

Ce fichier repose sur la syntaxe présentée dans le listing A.5. La syntaxe que nous avons définie permet aux auteurs de déclarer rapidement les sujets, scénarios et éléments de discours utilisables par le guide virtuel ainsi que les transitions et leur conditions d'application.

A partir de ces données, toutes les déclarations ainsi que les sons devant être joués pour les différents éléments de discours sont générés automatiquement. Un tel système est utilisable lors de la phase de prototypage afin de créer le contenu des éléments de discours constituant les explications du guide virtuel. Une fois cet ensemble de données stabilisé, il est néanmoins nécessaire d'effectuer un certain nombre d'ajustements notamment au niveau du choix du chemin au sein d'une explication (aléatoire par défaut) et de la synchronisation des sons avec des animations ponctuelles appropriées.

A.4 Définition du comportement

Un comportement est une machine à état permettant de contrôler les actions qu'effectue le guide virtuel dans l'environnement. Les actions sont déclenchées lors du passage de transitions valides au sein du comportement.

La définition d'un comportement passe par la description des différents états qui le compose ainsi que l'ensemble des transitions possibles entre ces états. Deux états particuliers doivent être définis obligatoirement :

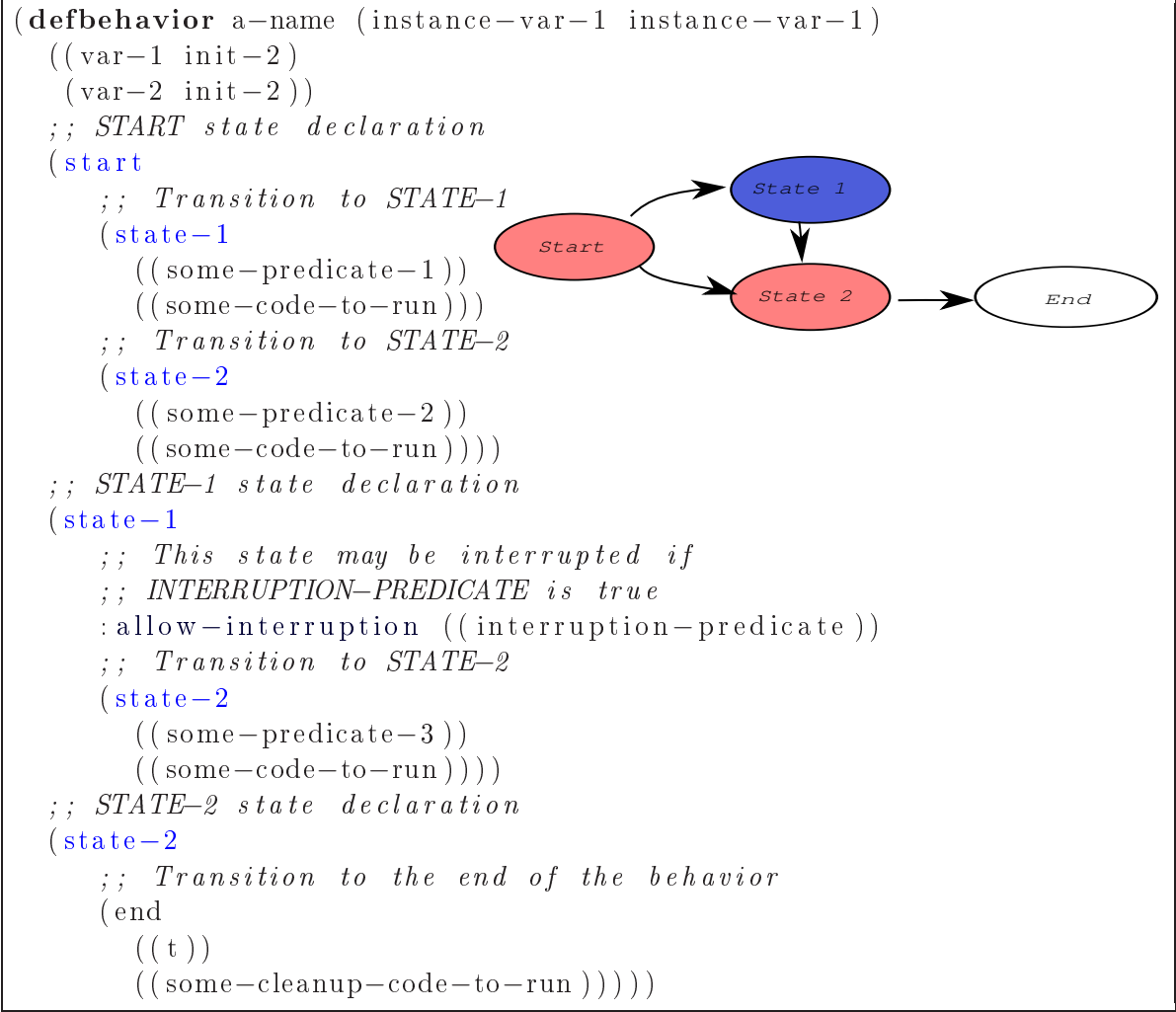
- ▷ **START** – L'état de départ du comportement permettant de s'assurer de la cohérence des différentes variables internes du guide virtuel.
- ▷ **END** – L'état de fin du comportement qui permet d'effectuer des actions spécifiques lorsque le comportement se termine comme par exemple envoyer un événement particulier.

L'état *END* est défini automatiquement et peut donc être référencé sans déclaration préalable. Ceci se justifie par le fait que cet état particulier ne permet aucune transition vers aucun autre état étant donné qu'il représente la fin d'un comportement. L'état *START* quant à lui doit être défini lors de la déclaration du comportement.

Le listing A.6 présente un exemple de déclaration d'un tel comportement. La déclaration est structurée comme suit. On déclare tout d'abord le nom du comportement qui permettra de le référencer ultérieurement ainsi que l'ensemble des variables locales au comportement. Ces variables peuvent se voir assigner une valeur par défaut qui sera affectée lorsque le comportement sera déclenché (dans l'état *START*). S'en suit la déclaration des différents états qui constituent le comportement. La déclaration d'un état doit comporter un identifiant permettant de référencer l'état en question ainsi que la liste de toutes les transitions partant de cet état. Chaque transition étant constituée d'un prédicat et d'une fonction devant être exécutée lorsque la transition est tirée.

Ainsi le code du listing A.6 permet de déclarer un comportement nommé *behavior-name* possédant deux variables locales (**var-1** et **var-2**) et les états et transitions suivants :

- ▷ L'état *START* qui doit être obligatoirement défini. Cet état possède deux transitions permettant d'atteindre les états *STATE-1* et *STATE-2* suivant respectivement les prédicats **some-predicate-1** et **some-predicate-2**. Chacune de ces transitions permet d'exécuter les actions appropriées lors du passage de l'état *START* à l'un des deux autres états.
- ▷ L'état *STATE-1* qui possède une transition vers l'état *STATE-2*. Cet état a la particularité d'être déclaré comme étant interruptible à condition que le prédicat **some-interruption-predicate** soit vrai.
- ▷ L'état *STATE-2* qui possède une transition vers l'état *END* qui a la particularité d'être toujours validée. Le passage de cette transition permet l'exécution d'actions spécifiques (**some-cleanup-code-to-run**) permettant de s'assurer de l'état du guide virtuel une fois le comportement terminé.



LISTING A.6 – Exemple de déclaration d'un comportement local et représentation graphique de celui-ci.

Comme nous l'avons précisé précédemment, l'état de départ du comportement doit obligatoirement être défini et doit porter l'identifiant *START*. Le comportement *END* est quant à lui défini automatiquement et peut être référencé directement.

Les variables *instance-var-1* et *instance-var-2* définissent des paramètres pouvant être utilisés lors de l'instanciation du comportement. Cette instanciation se fait de la façon suivante :

```
(make-behavior-a-name "value-1" 42)
```

La déclaration d'un état peut faire appel à ce que nous nommons des « modificateurs ». Dans l'exemple du listing A.6, l'état *STATE-1* utilise un tel modificateur permettant de rendre l'état interruptible (*:allow-interruption*). D'autres modificateurs existent tels que : *:with-action* ou *:composite* par exemple. Ceux-ci permettent respectivement d'associer des actions à un l'évaluation d'un état donné ou de déclarer un état comme étant composite (constitué d'un ensemble de sous-automates).

La définition de l'automate hiérarchique global se fait suivant la même syntaxe. La différence réside alors dans la présence d'états composites déclarés à l'aide du modificateur *:composite* et l'absence d'états interruptibles (modificateur *:allow-interruption*).

A.5 Définition et activation des experts

La sélection des informations est effectuée par un ensemble d'experts capables de donner leur point de vue sur une liste de sujets qui leur est fournie. La définition d'un expert se fait de la façon présentée dans le listing A.7.

```
(defexpert not-started-modifier
  (mapcar (lambda (a-subject)
            (if (not (subject-started-p a-subject))
                1
                0))
          subject-list))
```

LISTING A.7 – Exemple de déclaration d'un expert.

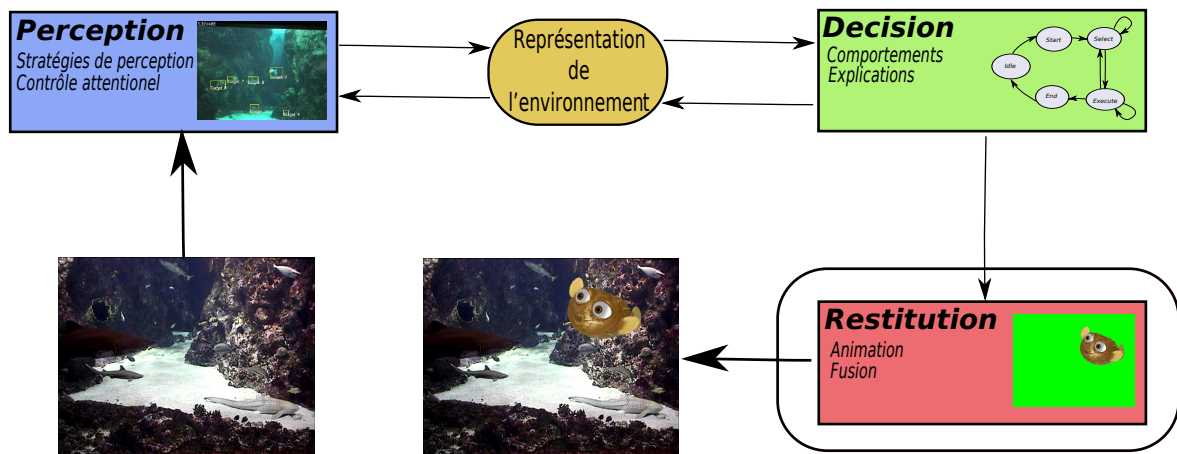
Cette définition nécessite le choix d'un identifiant qui sera utilisé lors de l'activation de l'expert en question. Cette activation se fait de la façon suivante :

```
(enable-subject-expert 'expert-name)
```

Cette activation ou la désactivation correspondante peut s'effectuer à n'importe quel moment durant l'exécution du module de prise de décision. Généralement celle-ci se fait lors de la phase d'initialisation.

Annexe B

Rendu et visualisation



B.1 Introduction

Le rendu de la représentation du guide virtuel est assuré par un module dédié réalisé par la société Polymorph Software¹. Cette représentation consiste en un poisson porc épïc à l'apparence amicale (figure B.1).

Ce module dédié est utilisé au travers d'un protocole de contrôle simple et prend en charge la gestion de la navigation du personnage virtuel. Lors de son exécution, le personnage est simplement représenté seul sur un fond monochrome permettant de faciliter la fusion avec le monde réel par le système de visualisation (figure B.2 en haut). Un programme client de test permet de se rendre compte des différentes fonctionnalités du système.

¹ <http://www.polymorph.fr>

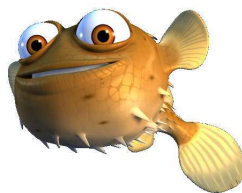


FIG. B.1 – Représentation du guide virtuel sous la forme d'un Diodon.

B.2 Protocole de contrôle

La contrôle du moteur de rendu s'effectue par une connexion réseau à l'aide d'un protocole texte simple orienté ligne de commande. Ce protocole est volontairement simple et de haut niveau centré sur le déclenchement d'animations et des commandes de navigation. Les différentes commandes existantes peuvent être classées en différentes catégories :

- ▷ les commandes générales ;
- ▷ les commandes de contrôle de l'état du guide virtuel ;
- ▷ les commandes de navigation ;
- ▷ les commandes d'animation et médias.

B.2.1 Commandes générales

QUIT

Permet de quitter l'application.

POSITION

Retourne la position courante du guide virtuel dans l'environnement 3D. La réponse est de la forme `POSITION <posx> <posy> <posz>`.

B.2.2 Contrôle de l'état du guide virtuel

MODE <unMode>

Description d'un mode d'action pour le guide parmi : poursuite, attente, monologue.

ETAT <unEtat>

Description de l'état interne du guide parmi : joie, ennui, peur.

B.2.3 Navigation

CIBLES <nbCibles>

Définit le nombre d'objets qui vont être envoyés pour la gestion de la navigation. En mode poursuite le premier objet envoyé est considéré comme l'objectif des déplacements du guide virtuel. Cette commande doit être suivi d'une succession de commandes **CIBLE** décrite ci-après

CIBLE <tag> <posX> <posY> <posZ> <rayon> <coeff>

Définition d'un objet entrant en compte dans le calcul de la trajectoire du guide virtuel. Le champ <tag> est pour l'instant ignoré. Les champs <rayon> et <coeff> permettent respectivement de spécifier la taille de la sphère englobante de la cible en question et de définir un poids particulier lors du calcul de trajectoire.

B.2.4 Animations et médias

JOUER ANIM <nom>

Joue l'animation spécifiée en paramètre. Cette animation « locale » se combine avec l'animation de déplacement existante. Pour l'instant les animations disponibles sont :

- ▷ Non
- ▷ Coucou
- ▷ Oui
- ▷ Parler

JOUER VIDEO <nom> <posX> <posY> <sizeX> <sizeY> <opacité>

Joue une vidéo aux coordonnées écran spécifiées. Les champs <sizeX> et <sizeY> permettent de spécifier la taille d'affichage de la vidéo à l'écran. Le champ <opacité> permet quant à lui de rendre cette affichage plus ou moins transparent.

JOUER VIDEO STOP

Arrete la lecture de la vidéo courante.

JOUER IMAGE <nom> <posX> <posY> <sizeX> <sizeY> <opacité>

Affiche l'image spécifiée aux coordonnées choisies. Les paramètres de taille et d'opacité ont ici la même signification que dans le cas de la lecture d'un document vidéo.

JOUER IMAGE STOP

Cacher l'image affichée

JOUER SON <nom>

Ajouter le son spécifié dans la file des sons à jouer.

JOUER SON STOP

Vide la file des sons à jour et termine la lecture en cours.

B.3 Gestion de la navigation

Le moteur de rendu et visualisation, en plus de se charger de l'affichage et des animations du guide virtuel, est responsable de la gestion de ses déplacements au sein de l'environnement.

Pour cela il fait appel à des algorithmes de *steering* mis en œuvre au sein de la bibliothèque OpenSteer². Ainsi du point de vue du module pédagogique la navigation dans l'environnement se fait simplement en transmettant la position et la taille des différents objets perçus dans l'environnement. A partir de ces informations, l'algorithme de steering est capable de calculer la trajectoire du personnage virtuel. Cet algorithme a été paramétré de tel sorte qu'il est impossible pour le guide virtuel de passer derrière un obstacle, éliminant ainsi les difficultés liées au calcul des occlusions qui reste un problème complexe dans le domaine de la réalité augmentée.

Les obstacles sont représentés sous la forme de sphères positionnées dans l'environnement virtuel (figure B.2). Il est possible de spécifier un coefficient d'importance aux différents obstacles afin de moduler leur influence sur le calcul de la trajectoire du guide virtuel.

² <http://opensteer.sourceforge.net>

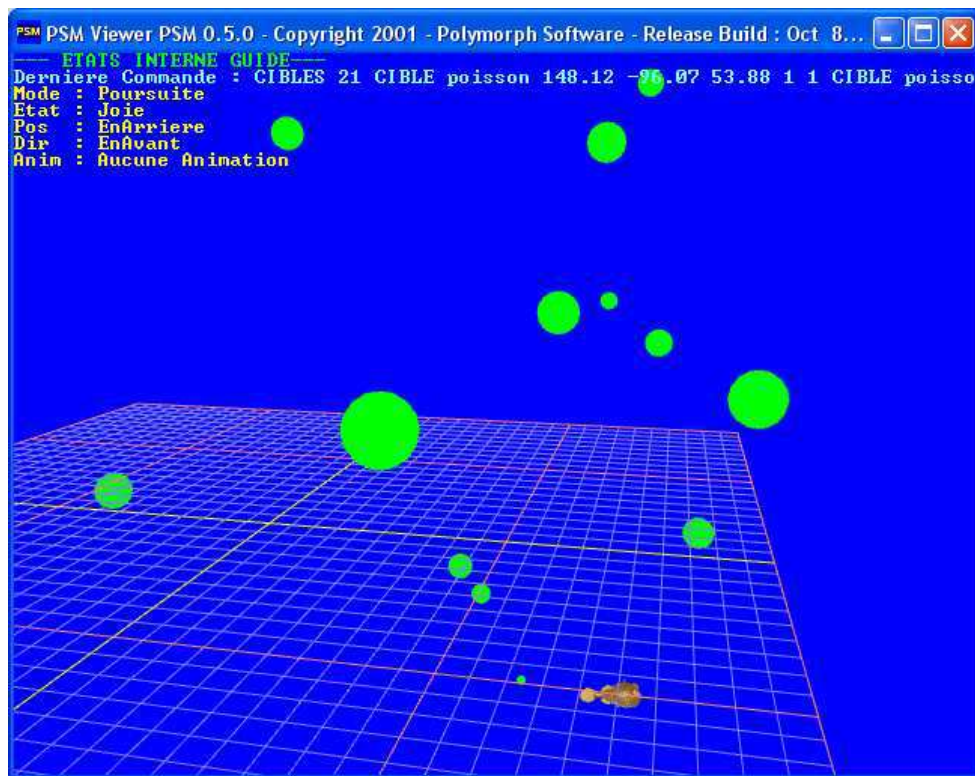
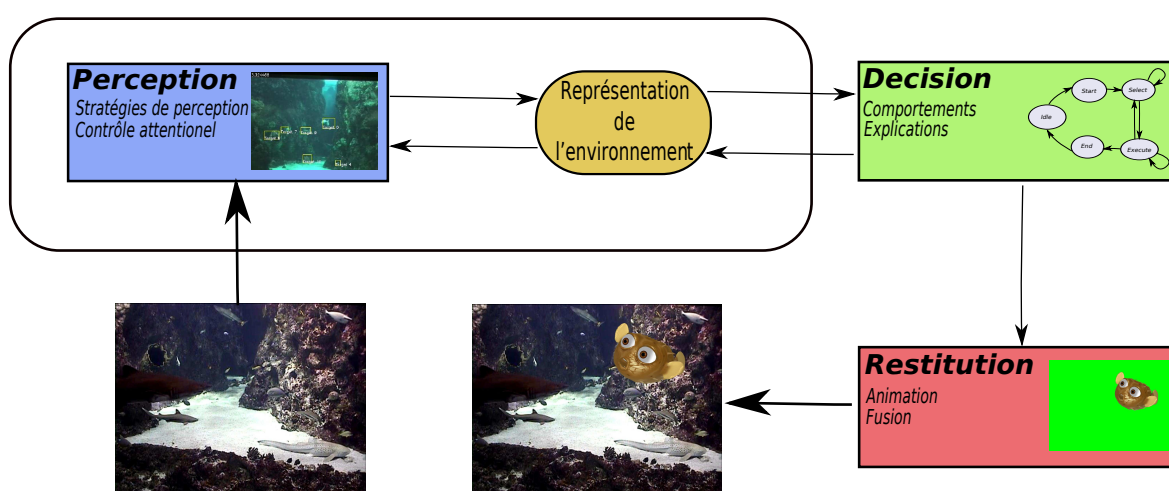


FIG. B.2 – Captures d'écran de la fenêtre du moteur de rendu avec l'affichage des cibles servant à la gestion de la navigation.

Annexe C

Implémentation du système de perception visuelle



C.1 Architecture logicielle

Le système de perception visuelle que nous avons présenté dans ce manuscrit est avant tout un système de vision par ordinateur. Il traite des images provenant d'un flux vidéo afin d'en extraire des informations qui permettront de construire une représentation de l'environnement. Ces traitements sont organisés en un ensemble de routines contrôlées par des comportements de perception. Dans le chapitre 5 nous avons parlé de l'exécution asynchrone et simultanée de ces différents comportement et routines. Cette exécution est supportée par une architecture logicielle dédiée qui offre la souplesse et la robustesse nécessaire à la mise en œuvre du module de perception.

Cette architecture repose sur l'utilisation d'un ensemble d'objets actifs offrant des services et capables de communiquer par envoi de message. Ceci a un impact direct sur les possibilités

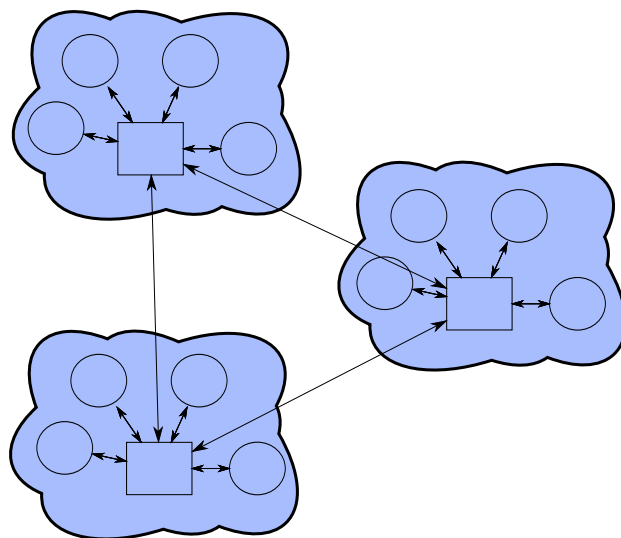


FIG. C.1 – Vue d'ensemble de l'architecture de communication et d'exécution d'objets actifs sur laquelle repose le système de perception visuel. Chaque « nuage » représente une plateforme d'exécution d'un ensemble d'objets actifs (les cercles). Chaque plateforme possède un module de communication (carré) se chargeant de la communication par message entre objets et de la recherche de service par un système d'annuaire.

d'exécution des différents composants et le fonctionnement du module de perception. Cela implique notamment l'absence d'une mémoire centralisée au profit d'une approche distribuée offrant des possibilités intéressantes du point de vue de l'exécution.

C.1.1 Un ensemble d'objets actifs qui coopèrent

L'architecture que nous proposons repose sur la notion d'*objet actif*. Un tel objet effectue un certain nombre de traitements et est capable de communiquer avec les autres objets afin de mener à bien ces traitements.

L'architecture logicielle permettant le support de l'exécution de ces objets actifs est présentée dans la figure C.1. Cette architecture présente la particularité de faciliter la distribution d'une application grâce à la notion de plateforme. Une plateforme se charge de l'exécution d'un ensemble d'objets actifs ainsi que des mécanismes permettant la communication entre ces objets (cf. C.1.2).

Les différentes tâches que peut effectuer un objet sont définies en terme de services enregistrés dans un annuaire centralisé pour chaque plateforme. Un service possède un nom permettant de l'identifier ainsi qu'une liste de propriétés permettant de le décrire précisément par rapport aux autres services équivalents qui possèdent le même nom. Pour effectuer un service, un objet actif particulier peut avoir besoin d'autres services qu'il est en mesure de localiser grâce aux différents annuaires. Les annuaires sont synchronisés entre les différentes plateformes pour des questions d'efficacité lors des requêtes. Un annuaire contient ainsi la liste

de tous les services existant parmi les différentes plateformes en cours d'exécution associée à la liste de tous les objets actifs fournissant ce service. Pour chaque fournisseur il existe une description détaillée du service particulier rendu à l'aide de propriétés (listing C.1).

```

Classification:
    provider:5
        levels = 0
        types = red
    provider:6
        levels = 0
        types = green
    provider:7
        levels = 0
        types = blue
UpdateInput:
    provider:0
        image-height = 240
        image-width = 320
        name = vision

```

LISTING C.1 – Exemples de services et de leurs propriétés. Ici deux services sont disponibles, l'un de classification et l'autre de mise à jour d'image à partir des caméras. Plusieurs objets actifs peuvent fournir le même service, un service particulier étant décrit par un ensemble de propriétés.

Une telle architecture nous offre une grande souplesse dans la mise en œuvre du module de perception (conception en terme de services simples), des possibilités intéressantes de distribution des calculs sur différents processeurs locaux ou distants¹ ainsi qu'une meilleure robustesse².

C.1.2 Exécution et communication

Au sein d'une même plateforme, les différents objets sont exécutés de façon séquentielle dans un ordre établi aléatoirement au début de chaque boucle d'exécution afin d'éviter des biais pouvant être introduit par l'ordre d'exécution.

L'exécution est considérée comme asynchrone, la communication entre objets actifs se faisant par un mécanisme d'envoi de messages. Lors de leur exécution certains objets ne font que réagir sur réception d'un message tandis que d'autres possèdent la capacité d'être pro-actifs et donc d'effectuer des tâches de leur propre initiative.

Un message contient les informations suivantes :

¹ La communication se fait de façon transparente au travers d'un réseau.

² Un objet actif ne disposant pas des services nécessaires à une tâche n'entraîne pas l'arrêt du système par exemple.

- ▷ un type ;
- ▷ un émetteur ;
- ▷ un destinataire ;
- ▷ un nom de service ;
- ▷ un ensemble de données spécifiques au message en question sous la forme d'une liste de champs `nom : valeur`.

Les différents objets sont identifiés de façon unique sur chaque plateforme et par une paire (`id_plateforme`, `id_objet`) pour la communication entre plateformes d'exécution. L'envoi de messages peut être de deux types : point à point ou diffusion.

```
Sender:    1@platform1:52135
Receiver:  2@platform2:42511
Type:      Reply
Service:   CurrentTimeStamp
Data:
    time=42
```

LISTING C.2 – Exemple de message

L'ordre des messages est conservé lors du passage d'une plateforme à une autre et en interne sur une même plateforme lors de l'envoi de plusieurs messages successifs à un objet actif. Il n'est par contre pas garanti que d'autres messages ne puissent s'intercaler dans cette succession.

Le type des messages n'est pas défini *a priori*. Néanmoins, deux types de message sont utilisés fréquemment et permettent la demande d'un service et la réception du résultat de cette demande : `Request` et `Reply`.

C.2 Des objets actifs au système de perception visuelle

Nous nous basons sur l'architecture logicielle que nous venons de présenter afin de mettre en œuvre le système de perception visuelle présenté en 5.5.

Deux questions sont ici à considérer :

- ▷ Quels objets ?
- ▷ Quels services ?

Ces deux questions se posent pour chacun des composants de notre architecture de perception visuelle.

C.2.1 Comportements et routines de perception

Tel que nous l'avons envisagé, chaque routine de perception correspond à un type d'objet actif au sein de l'architecture logicielle. La différence entre une routine de perception et un comportement s'effectue alors sur l'aspect pro-actif ou non d'un objet actif.

Nous avons mis en œuvre les objets actifs suivants :

- ▷ **vigilance, tracking, recognition** qui implémentent respectivement les comportements de perception *Vigilance*, *Suivi* et *Reconnaissance*.
- ▷ **classifier et feature** qui permettent de déterminer l'appartenance d'une entité visuelle à une classe d'objets particulière.

Ces différents objets actifs dépendent les uns des autres afin de construire une représentation de l'environnement exploitable par le module de prise de décision du guide virtuel. Ces dépendances sont définies par les services nécessaires à la réalisation d'une tâche particulière. Lors du lancement de l'application, chaque objet actif se charge de trouver des prestataires pour chacun des services dont il a besoin. Certains services sont nécessaires au bon fonctionnement d'un objet particulier. C'est le cas par exemple des services de traitement d'image (cf. C.2.2) pour les objets mettant en œuvre les comportements de *Vigilance* et *Suivi*. D'autres services au contraire ne sont pas nécessaires et c'est leur présence ou non qui va déterminer les possibilités d'un objet actif particulier. C'est le cas de l'objet **recognition** qui met en œuvre le comportement de *Reconnaissance*, les différents types qui peuvent être identifiés dépendent des classifieurs disponibles à l'exécution.

C.2.2 Traitement d'image

Le rôle du module de perception visuelle est de construire une représentation de l'environnement acquis par un ensemble de caméras vidéo. Pour cela il est nécessaire d'effectuer des traitements sur les images provenant de ces caméras. Ces traitements sont sous la responsabilité de l'objet actif **vision**. Celui-ci est capable de récupérer des images provenant des sources vidéo, d'appliquer des filtres sur ces images afin de créer des cartes de propriétés et enfin d'extraire des informations à partir de ces images ou de ces cartes.

Afin d'assurer la cohérence de ces informations avec l'état de l'environnement réel certains de ces traitements doivent être effectués en temps réel, c'est le cas par exemple des traitements relatifs au suivi des entités visuelles présentes en mémoire. C'est ici le facteur limitant du module de perception : seule une partie des traitements nécessaires à une représentation complète de l'environnement sont réalisables compte tenu de cette contrainte et d'une puissance de calcul limitée. L'objet actif chargé du traitement d'image implémente cette contrainte en offrant la possibilité de décider d'un nombre minimum d'images devant être traitées par seconde. En deçà de ce nombre, les traitements encore non effectués sont annulés et une erreur est envoyée aux objets ayant initié ces demandes de traitement.

Il est important de noter également qu'un seul objet de traitement d'image est présent au maximum par plateforme d'exécution. Les objets actifs s'exécutant de manière séquentielle sur une plateforme il n'y a aucun avantage à la présence de plusieurs d'entre eux. De plus ceci nous offre un meilleur contrôle du point de vue de la limite en terme de nombre d'images par seconde que nous venons d'évoquer.

C.2.3 Mémoire visuelle

Le système de perception visuelle étant mis en œuvre de façon distribuée, il n'existe pas de mémoire centralisée à proprement parler. La représentation de l'environnement se trouve répartie au sein des différents objets actifs, chacun ayant la charge de maintenir les propriétés qu'il est en mesure de calculer. Des messages sont diffusés lors de chaque mise à jour de ces informations afin d'assurer la cohérence globale de la représentation.

Un objet actif au sein du système complet possède l'autorité quant aux entités présentes en mémoire. C'est lui qui implémente les mécanismes d'oubli que nous avons décrits. Dans notre implémentation c'est l'objet **tracking** qui se charge de cette tâche.

Un objet actif spécifique a la charge de rassembler les informations de mémoire distribuée afin d'en extraire une représentation cohérente de l'environnement. Cet objet **controler** a également la charge de la communication avec le processus de prise de décision, offrant une interface simplifiée de communication avec l'extérieur du système centrée sur les mises à jour de la représentation et le contrôle des objets d'intérêt.

C.2.4 Notes sur la distribution

Ainsi que nous l'avons évoqué tout au long de cette annexe, l'architecture logicielle employée est par essence distribuée (objets actifs et communication par envoi de messages). Du point de vue de l'implémentation du système de perception visuelle cela offre certaines opportunités sur lesquelles nous nous attardons ici.

Tout d'abord cela permet d'exécuter différents ensembles de routines de perception au sein de différentes plateformes, chacune possédant son propre module de traitement d'image. Ceci permet par exemple d'offrir la puissance de calcul nécessaire aux comportements de suivi et de vigilance tandis que le comportement de reconnaissance s'exécute au sein d'un processus différent³.

Cela permet ensuite de dupliquer les unités de traitement afin d'en effectuer plusieurs en parallèle. Ceci peut s'avérer intéressant du point de vue du comportement de reconnaissance par exemple et des classifieurs dont il dépend. Multiplier les différents classifieurs peut permettre d'engager la reconnaissance de plusieurs cibles simultanément augmentant ainsi les performances potentielles du système. Ceci implique bien évidemment que les multiples unités

³ Ici nous considérons soit plusieurs machines communiquant au travers d'un réseau, soit une unique machine possédant plusieurs processeurs/cœurs.

de traitement soient prisent en compte par le comportement lui-même lorsqu'il déclenche les traitements dont il a besoin.

Reconnaissance et classification des espèces de poisson

D.1 Classifieurs et *features*

La reconnaissance et l'identification des entités visuelles présentes dans l'environnement reposent sur l'utilisation de classifieurs. Un tel classifieur est capable à partir d'un vecteur d'information comportant par exemple des propriétés visuelles ou géométriques de déterminer son appartenance ou non à une classe d'objets similaires.

Dans le système actuel nous avons implémenté les classifieurs suivants :

- ▷ **IdentityClassifier** qui renvoie simplement la première composante du vecteur d'information ;
- ▷ **RangeClassifier** qui se base sur un intervalle de valeurs afin de déterminer l'appartenance ou non d'un vecteur de données à une classe particulière ;
- ▷ **SVMClassifier** qui se base sur une implémentation des « machines à vecteur de support » (aussi appelés « séparateurs à vaste marge ») fourni au sein de la bibliothèque *svm-light*¹.

Il est possible d'ajouter d'autres types de classifieurs tels que des réseaux de neurones par exemple, de façon relativement transparente du point de vue du système de perception.

Les propriétés pouvant être utilisées par ces différents classifieurs sont également réifiées dans le système de perception ainsi qu'au niveau de l'architecture basée service sur laquelle son implémentation repose.

Dans son état actuel le système propose les propriétés suivantes :

¹ T. Joachims, Making large-Scale SVM Learning Practical. Advances in Kernel Methods - Support Vector Learning, B. Schölkopf and C. Burges and A. Smola (ed.), MIT-Press, 1999.

- ▷ EvolutionArea
- ▷ BoundingBoxRatio
- ▷ LargeVerticalStripes
- ▷ Color
- ▷ Ellipse
- ▷ Histogram
- ▷ Convexity
- ▷ Concavity
- ▷ Compacity
- ▷ ShapeDescriptor
- ▷ Speed

Chacune de ces propriétés retourne un vecteur de valeurs qui peut alors être intégré au vecteur d'information requis par le classifieur.

D.2 Reconnaissance et apprentissage

Dans le cadre du projet SIRENE, nous avons considéré l'identification de trois espèces de poissons distinctes :

- ▷ Platax ;
- ▷ Naso ;
- ▷ Requins.

Le processus d'apprentissage des différents classifieurs est semi-automatisé. Pour cela, le système de perception visuelle propose un mode de fonctionnement spécifique permettant d'extraire les données nécessaires à cet apprentissage. Dans ce mode de fonctionnement les différents comportements de perception (vigilance, suivi et reconnaissance) restent inchangés. La seule différence avec le mode de fonctionnement « normal » est la présence d'un « classifieur » spécial capable de demander le calcul de toutes les propriétés sur l'ensemble des entités présentes en mémoire. Ceci montre l'intérêt de l'implémentation souple à base d'objets actifs communicants et de services que nous avons choisie. Les données résultantes de l'exécution du système dans ce mode de fonctionnement sont ensuite stockées pour la seconde phase du processus d'apprentissage. Cette dernière est la seule phase manuelle du processus et consiste en la demande à l'utilisateur du type des différentes cibles provenant de la phase d'extraction de données. Pour cela il se voit présenter successivement les différentes entités ayant été

détectées en mode extraction et doivent déterminer son type interactivement ou simplement rejeter l'échantillon du processus d'apprentissage qui suit. Une fois cette phase terminée, l'utilisateur doit alors sélectionner les différents classifieurs et propriétés associées pour lesquels il souhaite effectuer l'apprentissage. La phase d'apprentissage à proprement parler est alors automatisée pour les classifieurs qui le permettent.

Pour la reconnaissance des espèces de poissons considérées dans le cadre du projet SIRENE, l'apprentissage a été effectué à partir d'environ 1000 échantillons au total (300 par classe et 100 faux échantillons). La résolution de ces échantillons varie de 20x20 à environ 150x60. Pour l'apprentissage d'une classe donnée, on utilise les 300 bons échantillons associés à 700 faux échantillons. Sur cet ensemble, les deux tiers sont utilisés pour l'apprentissage réel alors que le dernier tiers sert à tester le résultat de cet apprentissage.

Résumé

Classiquement la réalité augmentée consiste en l'annotation d'objets fixes pour un utilisateur en mouvement au sein d'un environnement réel. Le travail présenté dans cette thèse porte sur l'utilisation de la réalité augmentée dans le but de décrire des objets dynamiques dont le comportement n'est que peu prévisible. Nous nous intéressons tout particulièrement aux problèmes posés par la nature dynamique de l'environnement en ce qui concerne : 1°/ la *description du réel* (adapter les explications fournies par le guide virtuel à l'évolution de l'environnement) ; 2°/ la *perception du réel* (percevoir l'environnement à expliquer en temps réel à l'aide de caméras). La description du réel consiste en la génération d'un exposé par le guide virtuel. Cette génération repose sur deux points : des connaissances *a priori* sous la forme d'explications et un comportement décrit par un automate hiérarchique. Nous considérons la visite guidée comme l'évolution conjointe du comportement du guide virtuel et des explications qu'il fournit aux visiteurs. Une explication permet de décrire l'enchaînement d'éléments de discours sur un sujet donné à l'aide d'un graphe. Chacun de ces éléments décrit une unité de discours indivisible décrivant l'utilisation des différentes modalités (parole, gestes, expression, ...) sous la forme d'un script. L'exécution d'un graphe d'explication est effectuée par le comportement qui intègre la notion d'interruption. Lorsqu'un processus d'explication est interrompu, il est suspendu et le sujet courant de la visite guidée est réévalué. Cette réévaluation repose sur l'utilisation d'un ensemble d'experts votant pour les différentes explications disponibles selon un point de vue particulier. Ce vote se base sur le contexte courant de la visite guidée (historique, temps écoulé/restant, ...) et l'état de l'environnement réel. La perception consiste en la construction et la mise à jour d'une représentation de l'environnement. Ceci est effectué en temps réel par la coopération de différentes routines de perception. La complexité de l'environnement observé (quantité d'informations et variations des conditions d'éclairage) empêchent une analyse complète du flux vidéo. Nous proposons de surmonter ce problème par l'utilisation de stratégies de prise d'information adaptées. Ces stratégies de perception sont mises en oeuvre par certaines routines au travers du choix et du paramétrage des traitements qu'elles effectuent. Nous présentons un ensemble minimal de routines nécessaires à la construction d'une représentation de l'environnement exploitable dans le cadre de la description de cet environnement. Ce système repose sur la mise en oeuvre de trois stratégies de perception : la vigilance qui coordonne des traitements de détection dans le temps et dans l'espace ; le suivi qui se charge de mettre à jour les propriétés spatiales des entités existantes dans la représentation ; la reconnaissance dont le rôle est d'identifier ces entités. L'efficacité des stratégies de perception suppose une interaction entre la prise de décision (génération de l'exposé) et la perception (construction d'une représentation de l'environnement) de notre acteur virtuel autonome. Nous proposons de mettre en oeuvre cette interaction au travers de la représentation de l'environnement et des requêtes effectuées par le processus de prise de décision sur cette représentation. Nous avons mené des expérimentations afin mettre en évidence le fonctionnement des différents aspects de notre proposition et de la valider des conditions contrôlées. Ces travaux sont appliqués à un cas concret d'environnement réel dynamique complexe au sein du projet ANR SIRENE. Cette application met en évidence les questions liées à notre problématique et montre la pertinence de notre approche dans le cadre de la présentation d'un aquarium marin d'Océanopolis.

Mots clefs : réalité augmentée ; vision par ordinateur ; perception active ; acteur virtuel autonome.

Abstract

Augmented reality usually consists in the addition of virtual information to real static objects. Our proposal is to use augmented reality in order to annotate dynamic objects with non-predictable behaviors. We focus on problems related to the dynamic aspect of the environment : 1°/ *describing reality* (adapting the explanations provided by the virtual guide to the evolution of the environment) ; 2°/ *perceiving reality* (the real world using video cameras). The description of the reality is achieved through the generation of the virtual guide's discourse. This generation is based on two aspects : *a priori* knowledge represented as explanations and the guide's behavior modeled with a hierarchical state machine. We consider the guided tour as the joint evolution of this behavior and the explanations provided to the visitors. An explanation is a graph describing the organisation of a set of discourse elements for a specific subject. Each discourse element represents an atomic unit describing the use of different modalities (*e.g.* speech, gestures, facial expression) as a script. The execution of an explanation graph is done by the guide's behavior which manages interruptions. When an explanation process is interrupted, the current subject of the guided tour is re-evaluated. This evaluation is based on a set of experts, each one being able to vote for available explanations according to their point of view. This vote takes into consideration the current context of the guided tour (*e.g.* tour history, elapsed/remaining time) as well as the state of the real environment. The perception of the reality consists in the creation and update of an environment representation. This is achieved in real-time by multiple cooperating perception routines. The complexity of the observed environment (*e.g.* amount of available information, lighting conditions) prevents us from analyzing the entire video stream provided by the cameras. We propose to overcome this problem by the use of appropriate perception strategies. These strategies are implemented by specific perception routines which are able to select and parameterize computer vision treatments. We present a minimal set of perception routines required to build an environment representation usable by the virtual guide : vigilance schedules detection in space and time ; tracking updates spatial attributes of existing entities in the representation ; recognition identifies existing entities. The effectiveness of these perception strategies depends on the interaction between the decision process (discourse generation) and the perception (the building of an environment representation) of our autonomous virtual actor. We propose to implement this interaction using the environment representation and requests of the decision process on this representation. Experiments have been carried on in order to show the different aspects of our proposition and validate it in a controlled environment. This work has been applied to a real world dynamic environment in the ANR project SIRENE. This application exhibits problems related to our issue and demonstrates the relevance of our approach in the context of the guided tour for a marine aquarium in Océanopolis.

Keywords : augmented reality ; computer vision ; active perception ; autonomous virtual actor.